HIGH-STAKES READING ASSESSMENT AND ENGLISH ORAL LANGUAGE DEVELOPMENT: A STUDY OF THIRD GRADE ENGLISH LANGUAGE LEARNERS IN A TEXAS SCHOOL DISTRICT

A Dissertation

by

SANDRA T. ACOSTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Major Subject: Educational Psychology

HIGH-STAKES READING ASSESSMENT AND ENGLISH ORAL LANGUAGE DEVELOPMENT: A STUDY OF THIRD GRADE ENGLISH LANGUAGE LEARNERS IN A TEXAS SCHOOL DISTRICT

A Dissertation

by

SANDRA T. ACOSTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Rafael Lara-Alecio
Committee Members,	Patricia Goodson
	Beverly Irby
	Yolanda Padrón
	Fuhui Tong
Head of Department,	Victor L. Willson

May 2010

Major Subject: Educational Psychology

ABSTRACT

High-Stakes Reading Assessment and English Oral Language Development: A Study of Third Grade English Language Learners in a Texas School District. (May 2010)

Sandra T. Acosta, B.A., The University of Texas;

M.A., University of Wisconsin;

M.Ed., University of St. Thomas

Chair of Advisory Committee: Dr. Rafael Lara-Alecio

The purpose of this study was to examine: (a) the methodological quality of current research on English Language Learners (ELL) in the areas of high-stakes testing and oral language as a component skill of reading performance, (b) the association between oral language and reading performance in third grade Hispanic ELLS, and (c) the impact of instructional program model on ELLs' oral language development.

Two parallel systematic reviews were conducted searching CSA, Ebsco and Wilson electronic databases for empirical studies conducted in the U.S. and published in peer-reviewed journals in English. In study one, ELLs and high-stakes testing, eleven reviewed studies (N=11) met the inclusion/exclusion criteria; while in study two, ELLs and oral language, twenty-three reviewed studies (N=23) met the inclusion/exclusion criteria. Abstracted studies were evaluated using a 10-criteria matrix, and a methodological quality score was assigned.

The quantitative study used secondary longitudinal data from Project English Language and Literacy Acquisition (ELLA) to examine the relationship of ELLs' oral language and reading performance and the impact of instructional delivery model on oral language development. Project ELLA was conducted with Hispanic ELLs (*N*=185) over a 4-year period from kindergarten through third grade, and utilized a quasi-experimental design with four study conditions: *Structured English Immersion-Enhanced* (SEI-E, experimental), *Structured English Immersion-Typical* (SEI-T, control), *Transitional Bilingual Education-Enhanced* (TBE-E, experimental), and *Transitional Bilingual Education-Typical* (TBE-T, control). Oral language was operationalized using vocabulary and listening comprehension measures; reading achievement was operationalized using a state-wide reading assessment. Overall results indicated that English oral language accounted for approximately 22-30% of reading performance variability. Additionally, bilingual education instructional models (experimental and control) experienced the most growth in English verbal and listening comprehension skills compared to the growth rate of the SEI-T group. While students in SEI-E experienced greater increases than the SEI-T group in listening comprehension, there were no differences in growth rates for verbal skills between the two structured English immersion groups.

More research studies are needed on ELLs in secondary grades, a group poorly represented in the research studies reviewed. Also, more research is needed on the impact of test accommodations for ELLs on test validity evidence.

DEDICATION

To Efren

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Rafael Lara-Alecio and my committee members, Dr. Patricia Goodson, Dr. Beverly Irby, Dr. Yolanda Padrón, and Dr. Fuhui Tong, for their guidance throughout this process. I am especially grateful to Dr. Rafael Lara-Alecio and Dr. Beverly Irby for giving me access to the Project ELLA data.

There are others who have mentored me as a researcher and supported me with their friendship. I would like to thank Dr. Bruce Thompson and Dr. Oiman Kwok for their infinite patience and guidance in my journey through statistics. To Dr. Victor Willson, current EPSY Department Head, and to Dr. Michael Benz, former EPSY Department Head, thank you for your support. To the circle of study friends who are now valued colleagues and who mentored me through the dissertation process, I am forever grateful. I extend a heartfelt and special thanks to Heather Honoré, Hsien Yuan Hsu, Ebrar Yetkiner, and Cindy "the scientist" Warren. To Janie Kemp and student workers in the Bilingual Education Program, thanks for your help and understanding.

Por último quisiera agradecerle a Luisa López Grigera quien ha sido una inspiración para mí en sus trabajos como investigadora, en sus éxitos profesionales y en sus amplios conocimientos sobre la crítica literaria como mujer de letras. Para mis estudiantes, pasados y presentes, les agradezco su entusiasmo de aprender y tantas cosas que me enseñaron sobre el andar. Para concluir quisiera saludar a los recopiladores de la música y la voz popular argentina quienes en tantas madrugadas solitarias me

acompañaron con sus lamentos y dulces melodías, Mercedes Sosa, Atahualpa Yupanqui y Jorge Cafrune.

A mi familia, gracias por su paciencia y su apoyo moral durante estos años de estudio.

NOMENCLATURE

AYP	Adequate yearly progress
CFI	Comparative fit index
CREDE	Center for Research on Education, Diversity and Excellence
CSAP-E	Colorado Student Assessment Program in English
CSAP-S	Colorado Student Assessment Program in Spanish
EI	English immersion
EL2	English learned as a second language
ELL	English language learner
ELLA	English Language and Literacy Acquisition
EO	English only
ESL	English as a second language
ETAKS	Texas Assessment of Knowledge and Skills in English
FCAT	Florida Comprehensive Assessment Test
L1	First language, native language, or mother tongue
L2	Second language, language learned after first language
LAD	Language acquisition device
LC	Listening comprehension
LEP	Limited English proficient
LGM	Latent growth model
LPAC	Language Proficiency Assessment Committee

MQS	Methodological quality score
NAEP	National Assessment of Educational Progress
NCES	National Center for Education Statistics
NCLB	No Child Left Behind
NELS	National Education Longitudinal Study
OELA	Office of English Language Acquisition
PEIMS	Public Education Information Management System
PV	Picture vocabulary
RMSEA	Root mean square error of approximation
SAT-9	Stanford Achievement Test, Ninth Edition
SEI	Structured English Immersion
SEI-E	Structured English Immersion Enhanced/Experimental
SEI-T	Structured English Immersion Typical/Control
SEM	Structural equation modeling
SFUCD	San Francisco Unified School District
SRMR	Standardized root mean square residual
TAAS	Texas Assessment of Academic Skills
TAKS	Texas Assessment of Knowledge and Skills
TBE	Transitional Bilingual Education
TBE-E	Transitional Bilingual Education Enhanced/Experimental
TBE-T	Transitional Bilingual Education Typical/Control
TBOP	Transitional Bilingual Observational Protocol

TEA	Texas Education Agency
TEC	Texas Education Code
WLPB-R	Woodcock Language Proficiency Battery-Revised

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Two major trends will continue to shape U.S. education policy and public schools for the first half of the 21st century. Trend one is accountability models based on large-scale statewide assessments. Trend two is the demographic shifts in school populations. In the next section, I will provide a context for these trends and the potential impact on education.

Assessments and external accountability have become important policy tools under the No Child Left Behind Act of 2001 (NCLB; Lazear, 2006; Lee, 2008; Scheurich, Skrla, & Johnson, 2000; Valenzuela, 2005). As schools struggle to meet the achievement standards mandated by NCLB, student subpopulations and their learning needs are also changing. Over the past ten years, the ELL population has grown over 45% in Texas (Texas Education Agency [TEA], 2008a) and over 57% nationally (Ballantyne, Sanderman, D'Emilio, & McLaughlin, 2008). Given this scenario, reading achievement and achievement gaps based on subpopulation demographics are issues of national interest. Among Texas school children, English language learners (ELLs) comprise 15% of the 4.6 million Texas school children with Spanish speakers accounting for 92% of the over 120 languages spoken by Texas ELLs.

The purpose of my study was to examine: (a) the current research evidence on ELLs and high-stakes testing and oral language, and (b) the relationship between third

---

This dissertation follows the style of *American Educational Research Journal*.

grade Hispanic ELLs' English oral language development and their performance on the TAKS reading assessment in English. Toward this goal, I employed both qualitative and quantitative techniques. The qualitative phase of my study consisted of two parallel systematic reviews of the research literature, and the quantitative phase consisted of using structural equation modeling techniques to understand the relationship between oral language and reading performance.

Why examine oral language? Snow (1983) defined oral language as "all oral forms of communication, speaking, and listening" (p. 166). Oral language, one domain of communicative competence, is the ability of speakers to understand, to be understood, and to mediate understanding within a culture and within various contexts. Individuals can perform these linguistic functions because they understand the linguistic code and social rules of that culture (Saville-Troike, 1989). Thus, oral language within the context of linguistic competence and education has three important aspects in the language learning process: (a) as a social interaction, (b) as a cognitive tool to participate in the individual's own learning, and (c) as an entry to academic discourse (Saville-Troike, 1989; Snow, 1991).

Research evidence appears to support a relationship between oral language and certain reading skills in ELLs' primary and secondary language (Geva and Zadeh, 2006; Proctor, Carlo, August, & Snow, 2005; Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008). In a cross-sectional study of 1,531 Hispanic ELLs, researchers measured oral language and reading comprehension in both English and Spanish. Spanish oral language scores accounted for 10% of the variability in Spanish reading comprehension

scores. In comparison, English oral language scores accounted for 22% of the variability in English reading comprehension scores. All participants in the study were enrolled in bilingual transitional programs.[1] While adding to the body of research on biliteracy development, one of the limitations of cross-sectional research studies such as the study previously described is that no comparisons of the discrete features of oral language acquisition rates or trajectory are possible between grades (Miller et al., 2006).

Finally, two major research literature syntheses reported a paucity of research studies on oral language development in ELLs. Even more critical is the absence of longitudinal, experimental studies of literacy and academic oral English-language development among ELLs (Saunders & O'Brien, 2006; Slavin & Cheung, 2005; Tong et al., 2008). For this reason my study using secondary longitudinal data collected over a 4-year period in a quasi-experimental study of Hispanic ELLs and their English language acquisition rates addresses an important gap in the research literature on reading performance (i.e., reading literacy) and ELLs.

My dissertation used a journal article format and is organized into five chapters. In the first chapter (Chapter I), I introduce the study and create an organizational framework for the chapters that follow. Chapters II, III, and IV are formatted as journal articles.

---

[1] Bilingual transitional programs provide instruction in ELLs' home language as a temporary support for English language acquisition. The goal of these programs is to linguistically assimilate the ELL student into English-only instruction as soon as possible. Early-exit bilingual transitional programs allow home language support for two years, generally in kindergarten and first grade. Generally, late-exit bilingual transitional programs allow 40% instruction in the home language through the sixth grade (Baker, 2001).

The manuscripts for Chapters II and III present the results of two parallel systematic reviews of the literature: (a) high-stakes testing and ELLs presented in Chapter II, and (b) oral language and ELLs presented in Chapter III. The purpose of each study was to answer the following questions. *What are the methodological characteristics of the literature on high-stakes testing and ELLs published from 2001-2009? What are the methodological characteristics of the literature on oral language and ELLs published from 2004-2009?* A secondary aim of these studies was to categorize themes emerging from the research questions of the reviewed studies.

Systematic review procedures were employed based on a modified version of Garrard's (2007) Matrix Method for the health sciences and Slavin's (1986, 1995) best evidence approach for educational scientific research syntheses. I searched five electronic databases (ERIC, Ebsco-Academic Search Complete, PsycInfo, and Wilson Education Full Text) using search words related to English language learners, high-stakes testing, and oral language development. Studies that met the inclusion criteria were abstracted and their methodological quality was evaluated using a methodological assessment quality scale.

Study one (Chapter II) the systematic review of the literature on high-stakes testing and ELLs covers the period from 2001-2009. Eleven empirical studies (N=11) met the inclusion/exclusion criteria and were rated using the methodological quality assessment instrument. This is the only systematic review to my knowledge that has examined the methodological quality of empirical studies on high-stakes testing and

ELLs during the eight-year period since the passage of the No Child Left Behind Act of 2001 (NCLB).

Study two (Chapter III) the systematic review of the literature on oral language and ELLs is follow-up of the research literature synthesis on oral language and ELLs conducted by Genesse, Lindholm-Leary, Saunders, and Christian (2006) in *Educating English Language Learners: A Synthesis of Research Evidence*. My review covers the period from 2004-2009. In addition to the search of the electronic databases and purling, twenty-five referred journals were searched by hand. Twenty-three articles met the inclusion/exclusion criteria and were rated using the methodological quality assessment instrument. To my knowledge this is the first replicable, systematic review of the literature on oral language and ELLs since the publication of the synthesis conducted by Genesse et al. (2006).

The third manuscript (Chapter IV) is the quantitative study. In this study, I examine the relationship between third grade Hispanic ELLs' English oral language development and their performance on the Texas Assessment of Knowledge and Skills (TAKS) reading test in English. The TAKS is a state-wide assessment administered annually in Texas beginning in the third grade (For more detailed information on high-stakes testing, accountability, and ELLs in Texas, see Appendix A).The purpose of my study was to answer the following questions: *Does the English oral language proficiency trajectory of Hispanic ELLs, explain their performance on a third grade state-wide high-stakes reading test in English? Does ELLs' instructional program model assignment explain their English oral language trajectory?*

My study used secondary longitudinal data from Project English Language and Literacy Acquisition (Project ELLA), a federally funded grant from the U.S. Department of Education (# R305P030032) under the English Language Acquisition Evaluation Program (IES, 2008). In the Project ELLA study, researchers utilized a quasi-experimental[2] design (i.e., 2 x 2) with four conditions: structured English immersion (treatment and control) and a bilingual transitional model (i.e., one-way dual language model[3]; treatment and control). Data were collected from the Project ELLA participants, a cohort of Spanish-speaking students, beginning in kindergarten (fall, 2004) through the completion of third grade (spring, 2008; Unpublished Project ELLA Grant Performance U.S. Department of Education Report ED 524B 2007-2008). All research procedures for my study have been approved by the Institutional Review Board at Texas A&M University (see Appendix B).

---

[2] This study was a quasi experimental study because the schools rather than the student participants were randomly assigned as treatment or control. Schools selected to participate in the study had a school wide designation of SEI typical/control and TBE typical/control classrooms or of SEI experimental/treatment and TBE experimental/treatment classrooms. No participating school had a combination of treatment and control classrooms.

Student participants could not be randomly assigned to SEI or TBE classrooms because such assignments would be considered program placements. Under the Texas Education Code (TEC §29.063) Texas Administrative Code (TAC, Section 89.1220) Commissioner's rules (Chapter 189) the Language Proficiency Assessment Committee (LPAC) recommends program placement for ELL students, which is subsequently approved or not by the parent. For this reason, the researcher cannot randomly or otherwise assign an ELL student to an SEI or TBE classroom.

[3] Bilingual transitional program models use the primary language to move ELLs as quickly as possible into English-only instruction, usually within two years. Dual language programs aim to develop literacy in students' primary and secondary language, i.e., biliteracy. Two-way dual language program models include ELLs (language minority) and fluent English-speaking students (language majority) in the same classroom. One-way dual language programs denote one language group instructed bilingually, usually language minority students (Cloud, Genesee, & Hamayan, 2000; Ovando, Combs, & Collier, 2006; Skutnabb-Kangas & McCarty, 2008).

The two constructs for my study, oral language and reading performance were operationalized as follows. Oral language (i.e., verbal knowledge and listening comprehension) was operationalized using the Picture Vocabulary and Listening Comprehension subtests of the *Woodcock Language Proficiency Battery-Revised* (*WLPB-R)* English Form (Woodcock, 1991a); English reading performance was operationalized using the English TAKS (ETAKS) reading test for third grade. Results were analyzed using structural equation modeling techniques for fitting a latent growth model to longitudinal data.

Chapter V, the final chapter, contains a summary of the qualitative and quantitative manuscripts in my dissertation study. This chapter concludes with a final overview and discussion of (a) study findings and results, (b) the relationship of these studies to prior research on ELLs, (c) the implications for educational theory and practice, and (d) recommendations for future research and lessons learned.

CHAPTER II

HIGH-STAKES TESTING AND ENGLISH LANGUAGE LEARNERS: A SYSTEMATIC REVIEW OF THE RESEARCH EVIDENCE

Introduction

Over eight years have passed since the No Child Left Behind Act of 2001 (NCLB) was signed into law on January 8, 2002, by then-President George W. Bush. Educational reform under NCLB has reached the midpoint of the 13-year timeline for U.S. students to be proficient in reading and math. Under NCLB rules, 100% of students from kindergarten through twelfth grade in the 50 states, including Native Americans, and Puerto Rico would perform at grade level or above (i.e., be proficient) in reading and math by the school year 2013-2014. These performance outcomes were to be tracked utilizing benchmarks established by the states and Puerto Rico using large-scale state-wide assessments (i.e., high-stakes tests). Those schools not obtaining adequate yearly progress (AYP) established by their respective states and approved by the Department of Education would be identified and subject to *corrective action*. In tandem with NCLB's educational policy goal of reversing the academic achievement disparities or gaps between White students and other subpopulations, was the overarching social policy of educational parity for all students. English language learners (ELL) or limited English proficient (LEP) students comprised one of the NCLB target subpopulations and were a priority group, given the demographic profile of ELLs (Office of English Language Acquisition [OELA], 2008).

Students who speak a language other than English comprise 20% of the 50 million children ages 5-17 in U.S. schools. The majority are Spanish-speaking Hispanics of Mexican origin. Twenty-five percent of children who speak a language other than English at home speak English with difficulty (National Center for Education Statistics [NCES], 2009). This systematic review of the research literature synthesizes empirical research in one aspect of the education of ELLs in the United States from grades pre-kindergarten through the twelfth grade: academic achievement and assessment-based accountability.

Three empirical studies of the research literature on the topic of high-stakes testing and NCLB have been published in peer-reviewed journals from the years 2001-2009: Au (2007), a qualitative meta-synthesis published in *Educational Researcher*; Lee (2008), a meta-analysis published in *Review of Educational Research*; and Solórzano (2008), a comprehensive narrative review published in *Review of Educational Research*. While the research questions framing each review varied, a common factor in the three studies was change resulting from the associative relationship between test-driven external-accountability policies and student achievement.

In the first study, Au (2007) examined the impact of high-stakes testing policies on curricular control (i.e., curriculum and instructional delivery) in U.S. classrooms from grades K-12. The author reviewed 49 qualitative studies from the years 1992-2006. Employing a systematic review approach, Au described search terms, inclusion and exclusion criteria, and data collection and analysis in sufficient detail to permit replication. The author reported an overall interrater reliability of 89.4%. Study findings

generally suggested that high-stakes tests did change curricula through narrowing of curricula; fragmentation of content instruction into "small, individuated, and isolated test-size pieces" (Au, 2007; p. 262); and more-teacher centered instructional delivery. The author concluded that systemic curricular control emanating from a larger top-down policy design was supported by these findings.

The second study was a meta-analysis of 14 empirical studies published between 1994 and 2006. Lee (2008) analyzed effect size estimates reported by the reviewed studies of the relationship between accountability policies (high-stakes vs. low or nonaccountability states), the independent variable, and student academic performance outcomes, the dependent variable. The reviewed studies used national data from National Assessment of Educational Progress (NAEP) and National Education Longitudinal Study (NELS) test results for the years 1978-2002. Study findings suggested that high -stakes accountability policies exerted a slightly positive effect on student achievement but not on achievement disparities between ethnic subpopulations. It should be noted that the high-stakes testing and accountability policies discussed are pre-NCLB. With the exception of two reviewed studies, all remaining studies examined test scores from or before the year 2000.

Unlike the first two studies, Solórzano (2008) examined only ELLs and high-stakes testing. The author, employing a comprehensive narrative approach, reviewed 46 studies, over a 27-year period from 1981 through 2007. Data source materials (i.e., the reviewed studies) included empirical studies, commissioned reports, symposium findings, advisory board proceedings and government testimony. To be included in the

review, studies had to address one of three issues related to assessment of ELLs: determining academic achievement (e.g., norming and validity issues), language proficiency tests (e.g., defining English proficiency, academic predictive validity of proficiency tests), and fairness issues (e.g., accommodations, opportunity to learn). Solórzano's conceptual framework established validity as the principle construct and factor present in testing ELLs in a language other than their primary language.

My systematic review of the research literature contributes to research evidence on the topic of high-stakes testing and ELLs in three ways. First, to my knowledge no systematic, replicable review of the literature on high-stakes testing and ELLs has evaluated the methodological quality of studies examining high-stakes testing and ELLs. Second, while the studies conducted by Au (2007) and Lee (2008) examined the impact of high-stakes accountability policies and student achievement, their inclusion criteria and their findings did not address ELLs. Third, whereas the Solórzano (2008) study did examine ELLs and high-stakes testing, because of the holistic and comprehensive nature of this narrative review, it does not encapsulate the current state of the research literature regarding high-stakes testing and ELLs since the passage of NCLB.

Purpose

My study examines current research evidence on ELLs' English literacy development and their progress towards academic achievement parity with their native-English speaking peers. Proxy measures for academic achievement parity include external accountability assessments such as state-mandated testing instruments. These instruments benchmark adequate yearly progress (AYP) goals required by NCLB and

reported to the U.S. Department of Education. The *what* of my study are the resulting themes and issues derived from the research questions of the abstracted research studies, while the *how* is the validity and generalizability (quantitative research) or trustworthiness and transferability (qualitative research) of the research results or findings. In this study the yardstick for validity/trustworthiness is the methodological quality of these studies.

The literature on high-stakes testing and ELLs was systematically reviewed to answer this question.

1. What are the methodological characteristics of the literature on high-stakes testing and ELLs published in the years 2001-2009?

Conceptual Framework

High-stakes testing in education has been documented as early as the seventh century in the Kenju examination system of China, which utilized large-scale high-stakes tests for educational testing, for test-driven education, and for civil service exams (Suen & Yu, 2006). In the U.S., large-scale standardized high-stakes tests to assess student academic achievement and to evaluate instructional programs are the centerpiece of NCLB's accountability model. Faced with the rapidly changing demographics of American school children and the persistent achievement gap between White students and other subpopulations defined by ethnicity, economic disadvantage, and language, the federal government's role in education has evolved into a *hands on* (e.g. performance target, approval of state accountability plans), top-down management approach to enforce standards-based educational reform (Au, 2007; Popham, 2008). The missing

piece for ELLs in the external policy framework is a comprehensive language policy based on English language acquisition theory, pedagogy, and best practices to guide high-stakes testing of ELLs. Finally, ELLs' English language proficiency level is the *hidden factor* or phenomenon that impacts the validity and interpretation of English high-stakes testing results for ELLs.

While high-stakes testing has many interpretations, Solórzano offers a pithy alternative. High-stakes testing is "testing with major consequences for all involved" (2008; p. 263). For the purposes of this review high-stakes testing is defined as large-scale state-wide assessments aligned to state prescribed standards and employed by states as an accountability tool to measure the academic performance of students and to evaluate program effectiveness of schools and local education agencies (i.e., school districts). Thus, in the literature high stakes assessments are operationalized as state-wide assessments (e.g. Texas, Florida) of student achievement (e.g. reading, science).

Method

This systematic review of the literature consists of two parallel systematic reviews of empirical studies. Study one examines ELLs' performance and its relationship to high-stakes assessments and is presented in Chapter II. Study two, a follow-up study of an earlier research literature synthesis conducted by Genesse et al. (2006), examined ELLs' oral language and the relationship to literacy (study two is presented in Chapter III). My review was conducted using a modified version of Garrard's (2007) Matrix Method for the health sciences and Slavin's (1986, 1995) best evidence approach for educational scientific research syntheses. Two sources were

employed to locate scientific journal articles: electronic databases and purling. Once empirical studies addressing these relationships' were selected, articles were abstracted and analyzed using the Garrard matrix and coding sheets constructed for this systematic review. Modifications to the coding sheets in my study were informed by Honoré's (2008) systematic review of genetic risk and mate selection research literature in the health sciences. Research findings were synthesized and categorized according to the research questions found in the studies. These synthesized findings formed the basis for the present systematic review.

Inclusion and Exclusion Criteria

The empirical studies considered for inclusion in the systematic review of research evidence on high-stakes testing and ELLs were selected based on seven inclusion criteria. The articles had to: (a) be empirical studies (qualitative or quantitative); (b) be conducted in the United States; (c) be published in English; (d) be centered on some measurement of ELLs' cognitive, affective or linguistic development and high-states testing; (e) involve ELLs from prekindergarten through twelfth grades; (f) be published in the period from 2001 through 2009; and (g) be published in a referred journal.

The 2001 date was selected since this date corresponds to the passage of Public Law 107-100, NCLB Act of 2001, an act "to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind" (No Child Left Behind Act of 2001, p. 1425). NCLB mandates disaggregation of identified

subpopulations' scores when reporting student achievement performance. ELLs are designated as one of the statute-identified subpopulations (Popham, 2008).

Studies were excluded from the high-stakes review if they: (a) were reviews of the research literature, book reviews, commentaries, editorial articles, policy analyses, theoretical or framework proposals, unpublished manuscripts, and letters to the editor; (b) focused only on high-stakes testing and did not explicitly report a relationship between high-stakes testing and an underlying literacy foundation skill, (e.g. reading comprehension, verbal knowledge, etc.) or a relationship between high-stakes testing and cognitive, linguistic, or emotional development; (c) did not explicitly include ELLs in the sample and report finding for ELLs; and (d) were case studies of fewer than three ELLs.

Search Techniques

Four electronic databases were employed for article search and retrieval. These databases included ERIC (CSA), Academic Search Complete (Ebsco), PsycInfo (CSA), and Education Full Text (Wilson). Search strategies for high-stakes testing and ELLs employed variations of Boolean key terms and database descriptors. Search words were: *English language learners*, *language minority*, *high-stakes testing, accountability,* and variations of these words, e.g. *second language learners, second language learning, high-stakes tests.* Figure 1 presents the search words and Boolean terms used in the present systematic review.

Articles were screened and retrieved based on their titles and abstracts. Qualifying studies missed during the Boolean term search were supplemented by purling

the reference page of articles found from the searches and were added to the extant corpus of research studies. Articles gleaned from purling were also located via the searches of the electronic databases. Of the 162 studies retrieved and reviewed from the ELLs and high-stakes search, only 11 studies met the selection criteria (*N*= 11; see Appendix C-1 for a list of abstracted studies). All retrieved studies were assigned a number and inventoried in an Excel spreadsheet.

Data Abstraction and Analysis

The corpus of empirical studies included in the present systematic review of the research literature was generated in three steps: (a) a search of electronic databases in December 2008; (b) an examination of purled reference sections of retrieved articles; and (c) a second search of electronic databases in May 2009 to expand and locate any articles published between December 2008 and May 2009. In addition, the second search of electronic databases replicated the first search by combining the high-stakes and oral language searches (See Figure 2).

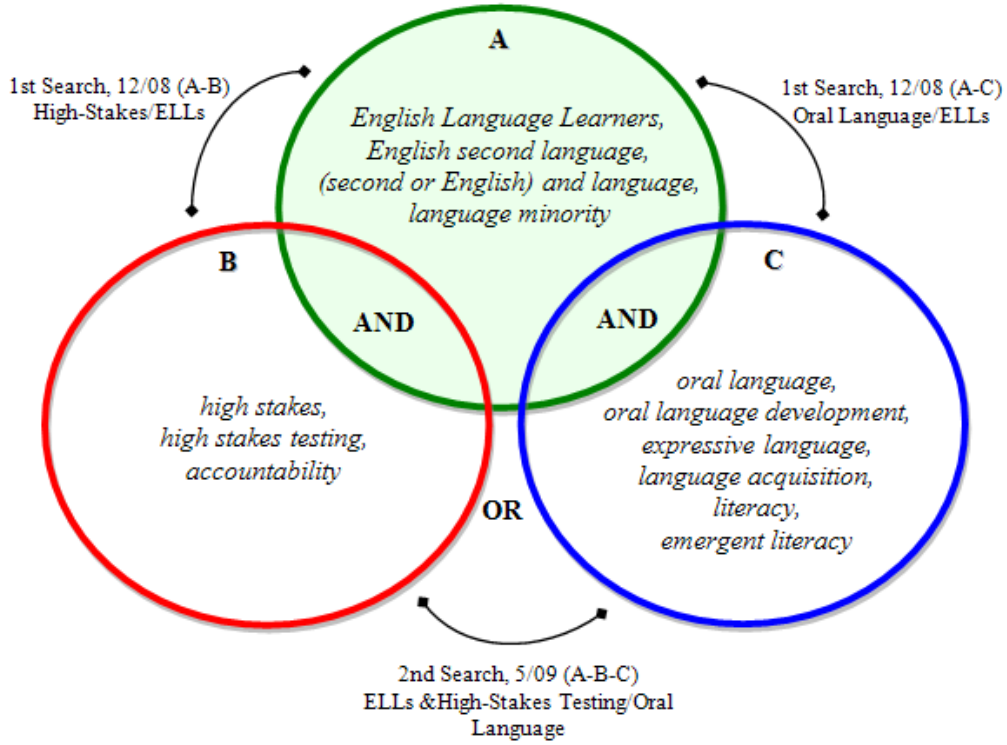

Figure 1. Search Word and Boolean Terms

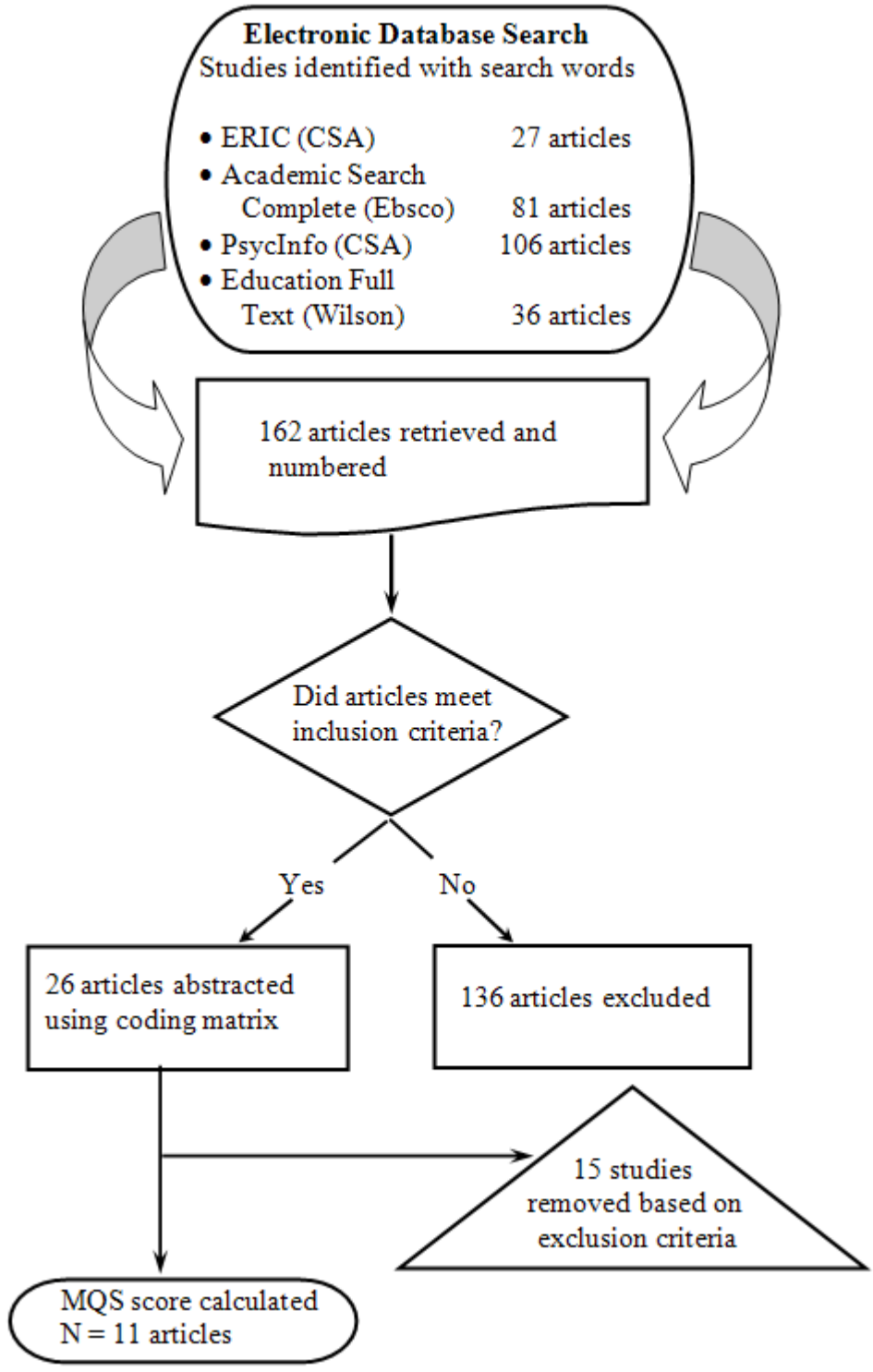

Figure 2. Methodological Steps: High-Stakes and ELLs Systematic Review

Methodological Quality Assessment

A content analysis and evaluation of the methodological quality of the sample articles was conducted using a matrix consisting of ten criteria: operationalization of the independent and/or dependent variables, theoretical or conceptual framework, instrumentation, research paradigm, research design, sampling, sample size, language groups, participants, and data analysis. Each criterion was assigned score points, which were then tallied. The methodological quality composite score (MQS) was the sum of the individual matrix criterion score points on the coding sheet (P. Goodson, personal communication, June 6, 2007; Goodson et al., 2006; Honoré, 2008). While the criterion were nominal categories, the MQS, a composite score, was intervally scaled. On this scale the higher the score the higher the quality of the study. The minimum possible MQS was 7 points with a maximum possible score of 25 points. The theoretical midpoint for the MQS was 16 points.

Each research study was treated as an observation, was scored, and received a MQS. Studies that met multiple elements within the same criterion, received the maximum score for that criterion. Criterion 7 "Sample Size" and criterion 10 "Data Analysis" included an additional scoring category for qualitative studies to ensure that qualitative studies were not more severely rated than quantitative studies (for a high-stakes test and ELLs sample abstraction sheet with MQS scoring rubric see Appendix E-1).

Table 1.

Summary of MQS Criteria Scores for High-Stakes Testing and ELLs Abstracted Studies (N=11)

Criteria	Qualitative Studies %	Quantitative Studies %	Mixed Methods Studies %	Total MQS Score Frequency %
C1- Operationalization of variables:				
No	33.3	42.9	00.0	**36.4**
Partial	33.3	00.0	00.0	**9.1**
Yes	33.3	**57.1**	**100.0**	***54.5***
C2- Theory:				
No	00.0	**57.1**	**100.0**	**45.5**
Yes	**100.0**	42.9	00.0	***54.5***
C3-Instrumentation:				
Reliability and validity unreported	33.3	**71.4**	00.0	***54.5***
Reliability and validity reported	**66.7**	28.6	**100.0**	**45.5**
C4-Research paradigm:				
Qualitative				**27.3**
Quantitative				***63.6***
Mixed Methods				**9.1**
C5-Research design:				
Idiographic/qualitative	**100.0**	00.0	00.0	**27.3**
Cross-Sectional	00.0	**42.9**	**100.0**	***36.4***
Cohort: Pre-Posttest	00.0	28.6	00.0	**18.2**
Cohort: ≥3 waves	00.0	28.6	00.0	**18.2**
C6-Sampling:				
Non-random	**100.0**	**57.1**	**100.0**	***72.7***
Non-random and comparison group	00.0	28.6	00.0	**18.2**
Random	00.0	14.3	00.0	**9.1**

Table 1. *(continued)*

Criteria	Qualitative Studies %	Quantitative Studies %	Mixed Methods Studies %	Total MQS Score Frequency %
C7-Sample size:				
Small - qual. = <30	33.3	00.0	00.0	**9.1**
Small - quan.= <30	00.0	00.0	00.0	**00.0**
Medium = >30 <300	**66.7**	14.3	**100.0**	**36.4**
Large = >300	00.0	**85.7**	00.0	***54.5***
C8-Language groups:				
ELLs	**66.7**	28.6	**100.0**	**45.5**
ELLs and EOs	33.3	**57.1**	00.0	**45.5**
ELLs-Dual Language Programs[b]	00.0	14.3	00.0	**9.1**
C9-Participants:				
Early childhood, P-2	33.3	00.0	00.0	**9.1**
Upper elementary, 4-6	00.0	14.3	00.0	**9.1**
Primary school, P-6	33.3	**71.4**	00.0	***54.5***
Middle school, 6-8	00.0	00.0	**100.0**	**9.1**
High school, 9-12	33.3	14.3[a]	00.0	**18.2**
C10-Data analysis:				
Univariate	00.0	14.3	**100.0**	**18.2**
Qualitative*	**100.0**	00.0	00.0	**27.3**
Bivariate	00.0	28.6	00.0	**18.2**
Multiple regression	00.0	14.3	00.0	**9.1**
Multivariate statistics (i.e, HLM, SEM)	00.0	**42.9**	00.0	**27.3**

Note. *Qualitative Analysis (e.g., content analysis; emergent themes analysis; grounded theory). See Appendix E-1 for abstraction sheet descriptors. Figures may not add to 100% due to rounding.

[a]The sample included grades 2-11 and was assigned the higher number.

[b]ELLs in dual language (DL) program models were included in the sample. DL models typically have EO (or fluent English speakers) and ELLs in the same cohort.

EO = English Only, fluent English-speaking (i.e., non-English language learner, who may or may speak or be fluent in other non-English languages); qual.= qualitative study; quan. = quantitative study

Table 2.

Evaluation Matrix: MQS Review Criteria Scores of High-Stakes Testing and ELLs Abstracted Studies

Reference		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	MQS total score
Qualitative studies (Ql):												
Booher-Jennings (2008)	Ql	1	2	1	1	1	1	1	1	1	2	12
Menken (2006)	Ql	2	2	2	1	1	1	1	1	3	2	16
Olson (2007)	Ql	0	2	2	1	1	1	1	1	1	2	12
Total qualitative studies	**M**	**1.00**	**2.00**	**1.67**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.67**	**2.00**	**13.33**
Range (12-16), Median=12	**SD**	**0.82**	**0.00**	**0.47**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.94**	**0.00**	**1.89**
Quantitative studies (Qn):												
Betts et al. (2009)	Qn	2	2	2	1	3	1	2	1	1	4	19
Escamilla et al. (2003)	Qn	0	0	1	1	1	1	2	1	1	1	9
Greene & Winters (2009)	Qn	0	0	1	1	2	1	2	1	1	4	13
Irby et al. (2007)	Qn	2	0	2	1	1	1	2	1	1	2	13
Lee et al. (2008)	Qn	0	0	1	1	2	3	2	1	1	4	15
Mahon (2006)	Qn	2	2	1	1	3	2	1	2	2	3	19
Tsang et al. (2008)	Qn	2	2	1	1	1	2	2	1	3	2	17
Total quantitative studies	**M**	**1.14**	**0.86**	**1.29**	**1.00**	**1.86**	**1.57**	**1.86**	**1.14**	**1.43**	**2.86**	**15.00**
Range (9-19), Median=15	**SD**	**0.99**	**0.99**	**0.45**	**0.00**	**0.83**	**0.73**	**0.35**	**0.35**	**0.73**	**1.12**	**3.38**

Table 2. *(continued)*

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	MQS total score
Mixed methods studies (MM):												
Shyyan et al. (2008)	MM	2	0	2	2	1	1	1	1	3	1	14
Total abstracted studies	**M**	**1.18**	**1.09**	**1.45**	**1.09**	**1.55**	**1.36**	**1.55**	**1.09**	**1.64**	**2.45**	**14.45**
Range (9-19) Median=14	**SD**	**0.98**	**1.04**	**0.52**	**0.30**	**0.82**	**0.67**	**0.52**	**0.30**	**0.92**	**1.13**	**3.11**

Note. See Appendix E-1 for abstraction sheet criterion descriptors.
Ql = Qualitative; Qn = Quantitative; MM = Mixed Methods. Skewness= 0.04

Results

The search of empirical journal articles published in peer-reviewed journals between 2001 and 2009 was conducted in two phases (December 2008 and May 2009). This dual-phase research yielded 250 articles related to high-stakes testing and ELLs. Article abstracts were scanned resulting in 162 articles that were retrieved and numbered.

The final sample (*N*=11) was selected through a two-tier process. The first tier selection process, which yielded 26 articles, consisted of reviewing each article and afterwards completing a coding sheet of inclusion criteria. The second review to ensure that all articles met both the inclusion and exclusion criteria yielded 11 articles (see Appendix E-1for abstraction sheets). Articles were excluded when they did not meet the inclusion criteria (e.g., were not empirical studies, did not relate directly to high-stakes testing and academic achievement for ELLs, focused on populations other than pre-kindergarten through twelfth grade students, or were conducted outside the U.S.).

Studies' Characteristics

Of the 11 reviewed studies, almost one third (27.3%) were published in 2008. No reviewed studies were published from the years 2001 through 2002. The remaining studies were evenly distributed (*n*=2 per year) over the years 2004, 2005, 2007, and 2009, excepting 2003 (*n*=1). Studies appeared in various U.S. professional journals, with the exception of one Texas study published in *British Journal of Sociology of Education,* and represented diverse disciplines in education: school psychology, linguistics, and economics. Three were published in one ELL-specific journal, *Bilingual Research*

Journal (BRJ), one article in 2002 and two articles in the same issue published in 2006. *BRJ* is an official publication of the National Association for Bilingual Education (NABE).

Authors generally identified the high-stakes tests employed in their studies (*n*=9). State-wide assessments represented six states: California (Stanford Achievement Test-Ninth Edition), Colorado (Colorado Student Assessment Program), Florida (Florida Comprehensive Assessment Test), Minnesota (Minnesota Comprehensive Assessment), New York (New York State Regents Exam), and Texas (Texas Assessment of Academic Skills and Texas Assessment of Knowledge and Skills). Of the content areas measured, reading test scores were the most commonly analyzed (63.6%), either alone (*n*=5), with math tests (*n*=1), or with writing tests (*n*=1).

Studies' Methodological Quality

Using the MQS protocol (see Appendix E-1) adapted from Honoré (2008), each study from the high-stakes testing and ELL sample (*N*=11) was evaluated and assigned a methodological quality score (MQS). The MQS is a composite of the 10 criteria scores. Table 1 presents the frequency distribution summary of the MQS criteria. MQS for reviewed studies ranged from 9 points to 19 points for the current sample (Table 2). Additionally, studies were grouped by research paradigm (i.e., qualitative, quantitative, mixed methods) with MQS scores reported as subtotals. The mean total MQS for all reviewed scores was 14.45 (SD=3.11). Variability was somewhat less in the qualitative studies (*n*=3) than the quantitative studies (*n*=7) as indicated by the larger range and standard deviation of the quantitative studies (see Table 2). The mean total value and the

median total value were almost identical (median=14) with a slightly positive normal distribution (skewness=.04) of the scores. Generally studies' MQS were below (63.6%; *n*=7) the theoretical midpoint of 16 points.

Most reviewed studies were quantitative (*n*=7), with qualitative studies (*n*=3) representing about one third of studies. One study utilized a mixed methods approach (Shyyan, Thurlow, & Liu, 2008). Although not self-reported as mixed methods by the authors, the study was labeled mixed methods because of companion qualitative and quantitative studies or phases. Mixed methods and the quantitative studies generally employed cross-sectional research designs (*n*=4) while the other quantitative studies used either pretest-posttest (*n*=2) or longitudinal[4] designs (*n*=2).

Sample sizes were generally medium (over 30 and under 300 participants) or large (over 300 participants). All studies but one utilized non-randomized sampling (*n*=10). While one third of the samples (*n*=4) had less than 100 participants, studies with large samples (63.6%) ranged from 300 to 123,347 students. Of these large sample studies, three used secondary data provided by state education agencies. Third graders were the most commonly represented group (71.4%) in the samples while secondary students were the least represented (i.e., middle school, *n*=1; high school, *n*=1). A third study used data from primary and secondary schools (grades 2-11). The qualitative and mixed methods studies (*n*=4) demonstrated the most variation, drawing samples from primary, middle school, and high schools.

[4] Longitudinal in the present systematic review uses Singer and Willett's (2003) definition of a longitudinal design as having three or more waves.

Overall, the language groups represented were evenly divided between ELL only and ELL and English-Only (EO) or fluent English-speaking participants. Escamilla, Mahon, Riley-Bernal, and Rutledge (2003) included participants from dual language classes[5] in their sample. Non-English primary language groups were generally Spanish-speaking ELLs (81.8%; *n*=8). Other primary language groups represented were Hmong, Chinese[6], Haitian Creole, and Somali. Excepting the Hmong-speaking ELL study, these studies also included Spanish-speaking ELLs (*n*=3) in their sample. In two studies (Betts, Bolt, Decker, Muyskens, & Marston, 2009; Tsang, Katz, & Stack, 2008), Spanish-speaking ELLs were the comparison group.

Data collection techniques generally consisted of standardized tests (*n*=7). Other data collection instruments were interviews (*n*=4), observations (*n*=4), and surveys (*n*=2). Lee, Maerten-Rivera, Penfield, LeRoy, and Secada (2008) employed a researcher developed science assessment in a pretest-posttest design. Fewer than half (*n*=5) of the reviewed studies reported validity and reliability statistics. Tsang et al. (2008), in addressing evidence of reliability, reported that the secondary data in their study did not have item scores; therefore, reliability could not be calculated. In contrast, more than half (*n*=6) of the reviewed studies operationalized variables or phenomena. Qualitative studies demonstrated more variation than qualitative studies with one qualitative study

[5] Dual language, also known as two way or two way immersion, refers to bilingual program models in the US that typically have grade level cohorts of language majority (L1= English) and language minority (L2=English) whose goal is to produce biliterate and relatively balanced bilingual individuals. In this case balanced bilinguals mean approaching fluency in both languages across contexts (Baker, 2001).

[6] Chinese was a generic language term used by the authors. The majority of the Chinese-speaking participants were Cantonese-speaking.

represented in each of the operationalized criteria categories. Quantitative studies either operationalized the variable or not. However, studies that provided operational definitions did not necessarily employ the word operationalize or its derivative. Only two studies used the term *operationalize* in their reports (Mahon, 2006; Tsang et al., 2008).

Theories explicating constructs and phenomenon were generally explicit for qualitative studies (*n*=3) and implicit for quantitative (*n*=4) and mixed methods studies. For studies without theoretical frameworks, linkage to theory was established utilizing the literature review. Of those studies explicitly discussing theory (*n*=6), more than 75% (*n*=4) cited Cummins' theories (i.e., common underlying proficiency, basic interpersonal communicative skills and cognitive academic language proficiency).

Statistical techniques for data analysis tended to be multiple regression (57.2%) in the quantitative studies with three studies reporting specific approaches (i.e., latent growth curve, logit, and hierarchical linear modeling). The remaining quantitative studies used either univariate (*n*=1) or bivariate analyses (*n*=2). The mixed methods study employed univariate statistics.

Studies' Empirical Findings

The reviewed studies contained a total of 109 findings with a mean of 5.03 findings per study. ELLs' academic achievement, measured by state-wide high-stakes assessments, was a dependent variable in all of the quantitative studies (*n*=7). In three studies, high-stakes assessment scores were also the independent variables. Independent or predictor variables for ELLs' academic achievement were acculturation (i.e., time in

the U.S.; *n*=1), LEP/ELL status (*n*=2), language of test (i.e., Spanish or English; *n*=1), reading achievement (*n*=2), and grade-level (*n*=1). Phenomena examined in the qualitative and mixed methods studies were achievement ideology (*n*=1) and instructional pedagogy (*n*=3).

A preliminary review of the findings produced three broad categories of the research questions found in the studies: studies examining English language (EL2) acquisition, studies examining second language acquisition pedagogy, and studies examining accountability-driven policies. Adopting a constant comparison methodology, three meta-themes emerged from the research question categories that encapsulated and provided cohesion to study findings. The three meta-themes are the following: (a) state-wide single assessment accountability, (b) English L2 literacy models and trajectories, and (c) narrowing of curricula and pedagogy.

Discussion

This review contributes to research evidence on ELLs in two ways. First, the review focuses on research published since the passage of NCLB and provides an overview of the impact of accountability assessment on ELLs' academic achievement. Secondly, this systematic review of the literature, focusing on the methodological quality of the corpus of studies, permits a replicable, critical evaluation of the literature that is absent in nonsystematic or narrative reviews.

The overall studies' methodological characteristics indicated the following. Third grade was the most commonly represented grade; studies were generally non-

randomized; and quantitative studies outnumbered by half qualitative studies. Finally, most studies fell below the theoretical midpoint for methodological quality.

A secondary aim of my study was to examine themes resulting from patterns identified in reviewed studies' research questions. Study findings were categorized by the research question. From these categories emerged three meta-themes. These meta-themes were: a) state-wide single assessment accountability, b) English L2 literacy models and trajectories, and c) narrowing of curriculum and pedagogy.

Meta-Themes

Meta-theme one: State-wide single assessment accountability. Arguably accountability is the most controversial aspect of large-scale high-stakes testing. Three studies examined the impact of state accountability policies. Escamilla et al. (2003) examined the impact of Colorado's accountability policy on ELLs and schools with large Hispanic ELL populations. Colorado policy at that time required the reporting of Colorado's state-wide assessment scores for English assessments but not for Spanish assessments, (i.e., the Colorado Student Assessment Program in English, CSAP-E; Colorado Student Assessment Program in Spanish, CSAP-S). Under state policy, ELLs' also received a 3-year exemption from taking the CSAP-E. Evidence from the authors' study demonstrated that third grade Hispanic students who took the CSAP-S scored higher than Hispanic students who took the CSAP-E but not higher than the general population of third graders taking the CSAP-E. Study findings also confirmed a negative relationship between schools with high ELL populations and these schools' performance ratings or report cards, even with the 3-year CSAP-E exemption policy. Additionally,

only the Escamilla et al. study included ELLs from dual language program models in their sample. From a methodological perspective, their report suffered from lack of clearly articulated theoretical linkages and operationalization of variables. However, one strength was the historical contextualization of Colorado's accountability policies and state legislators who supported English-Only language policies.

Student retention policies were one outcome of single assessment accountability, and presented the second issue. Two studies examined these phenomena from different perspectives. The first utilized an external, systemic perspective in evaluating a retention policy's impact on student academic performance. Greene and Winters (2009) analyzed third grade test scores of students scoring below promotion benchmarks on the Florida Comprehensive Assessment Test (FCAT). Their findings revealed that non-ELL Hispanics and Blacks were more likely to be retained than Whites under Florida's single assessment retention policy. Limited English Proficient (LEP)/ELLs had exemption status and therefore could be promoted without passing the FCAT. Achievement gains were calculated using the FCAT scores of retained and exempted students, including ELLs. FCAT scores were compared for two years following the baseline year in which the retention policy was first implemented for third grade. Study findings demonstrated that retainees' achievement gains continued to be higher than exempted students' gains. However, it should be noted that this comparison was based on FCAT scores from different grade level FCAT tests.

Booher-Jennings' (2008) qualitative study of achievement ideology provided the second perspective of single assessment accountability. This study examined the

socialization effect of motivating students to pass high-stakes tests from an internal, local perspective (i.e., a largely Hispanic school in Texas). Third grade students' perceptions of achievement were captured utilizing excerpts from student interviews about passing or failing the Texas Assessment of Knowledge and Skills (TAKS) test. Study findings suggested that gender-based differences in students' perceptions about achievement were related to internalized messages from teachers and administrators, based on these educators' views of girls' and boys' behavior at school. Thus, the author characterized girls' perceptions as related to self-esteem, whereas boys' perceptions were related to self-discipline. For example, girls received the message *to do their best work* whereas boys *to work harder*. Booher-Jennings noted that some boys began to express doubts about the fairness of the single assessment retention policy when they were confronted with the dissonance of failing the test after making their best effort. The author's failure to operationalize achievement ideology is problematic. The interview format is succinctly described as semi-structured and open-ended with no information about the interview protocol items or the theory and procedures used to frame the data collected.

Meta-theme two: English L2 literacy models and trajectories. These research studies examined content area literacies from three perspectives: (a) language of instruction time allotment in bilingual education program models (Irby, Tong, Lara-Alecio, Meyer, & Rodriguez, 2007), (b) instructional strategies (Lee et al., 2008; Shyyan

et al., 2008), and c) linguistic bias[7] and content assessment (Betts et al., 2009; Mahon, 2006; Tsang et al., 2008). Irby et al. asked bilingual education teachers to estimate the amount of time (*language allotment*) they spoke in Spanish or English during the science or social studies reading segment. The authors compared teachers' perception of language of instruction (LOI) to observed LOI. They also examined the relationship of LOI and language of the test (LOT; i.e. Spanish or English) to ELLs' performance on the Texas Assessment of Academic Skills (TAAS). Study findings suggested that bilingual education teachers severely miscalculated language distribution in their classrooms. Teachers tended to underestimate the frequency of Spanish while overestimating the frequency of English. Teacher observations were measured using the Transitional Bilingual Observational Protocol (TBOP; Lara-Alecio & Parker, 1994). The authors described the TBOP in detail and reported interrater reliability coefficients. Evidence also supported the associative relationship of TAAS performance, LOI, and LOT.

Two studies examined instructional strategies for ELLs. In the first, an experimental study, Lee et al. (2008) reported results for their randomized study of science instructional strategies for ELLs. Treatment and control classrooms were assigned based on teachers who received training in science instructional strategies for ELLs (treatment) and teachers who did not receive training (control). Using a researcher-developed science assessment in a pretest-posttest design, evidence supported science achievement gains overall for third grade students in the treatment classrooms. However,

[7] In this study linguistic bias refers to variability that is not due to random error but to the English proficiency level of a second language learner who is acquiring English (García, McKoon, & August, 2006).

when ELL and retainee subgroup scores were disaggregated from total scores, no treatment effect was noted.

Shyyan et al. (2008) compared how teachers and Hmong-speaking ELLs with disabilities ranked instructional strategies for reading, math, and science. Strategies included in the survey were selected from teacher focus group findings. Survey results revealed that teachers and students generally did not agree in their ranking of important instructional strategies. In ranking the importance of the three content areas, teachers and ELLs ranked science as least important. However, students ranked math as most important while teachers ranked reading as most important. The authors reported that all research participants were allowed opportunities to discuss the instructional strategies ranking process. Unfortunately no excerpts from these discussions, which would have contextualized the study, were included in their report.

Linguistic bias and English language acquisition trajectories were addressed in three studies. Tsang et al. (2008) analyzed secondary data from the San Francisco Unified School District (SFUCD). Their dataset focused on Spanish- and Chinese-speaking (the majority, Cantonese-speaking) ELLs and former ELLs from the 2nd through eleventh grades. The authors compared Stanford Achievement Test, Ninth Edition (SAT-9) reading and math scores of the two SFUCD ELL groups with SAT-9 national sample scores. They reported that the correlation difference between national sample scores and Chinese-speaking ELL scores converged at five years and after six years for Spanish-speaking ELL group. Because this study reported 2-11, they received the higher MQS for high school.

In the second study, Betts et al. (2009) examined the associative relationship between academic achievement and acculturation, (i.e., time in the U.S. and oral reading fluency), utilizing a sample of third grade Spanish- and Somali-speaking ELLs. Evidence supported initial oral reading fluency test levels and time in the U.S. as predictors of reading achievement, as measured by the Minnesota Comprehensive Assessment (MCA) third grade.

Finally, Mahon (2006) examined language proficiency and literacy performance in a study of fourth and fifth grade Spanish-speaking ELLs in Colorado. The author found a high positive correlation between the Woodcock Muñoz Language Survey (WMLS), a language proficiency test, and the Colorado Student Assessment Program (CSAP). Study findings revealed an achievement gap between ELL fifth graders in 2- and 3-year cohorts and all Colorado fifth graders. In the 2-year cohort, the reading achievement gap decreased by .12 of a standard deviation and the writing achievement gap decreased by .17 of a standard deviation. Similarly, in the 3-year cohort, the reading achievement gap decreased by .31 of a standard deviation. In the same study, the author also examined the crosslinguistic transfer of Spanish to English for ELLs by measuring the CSAP-S test scores for reading and writing and students' CSAP-E test scores. Evidence indicated that English and Spanish language CSAP tests were positively correlated for both content areas.

Meta-theme three: Narrowing of curricula and pedagogy. While narrowing of the curriculum is sometimes characterized as *teaching to the test*, it can also result in instructional changes that are less favorable to ELLs: instruction that is less student-

centered and more teacher-directed; pedagogy and instructional practices more closely aligned with English language arts classes (i.e., for monolingual English students); and instructional design incorporating fewer evidence-based instructional strategies for ELLs. Findings from studies conducted by Olsen (2007) and Menken (2006) supported the hypothesis that curricula were shrinking in the ELL classrooms observed. Findings from both studies indicated gravitation of instructional practices from English as a second language methodology towards English language arts methodology. Although study samples were drawn from different populations, primary (2nd grade) and high school, both authors reported more teacher-centered and less student- centered pedagogy. Additionally, in the case study of two bilingual classrooms, Olsen observed *hyper-scaffolded* instructional discourse, which the author described as a strategy to control student participation and talk. Olsen concluded that these teacher questioning techniques and discourse patterns restricted student and teacher feedback mechanisms, including accurate assessments of students' knowledge by the teacher. Interview data confirmed primary and secondary educators' commitment to primary language (L1) instruction (i.e., Spanish). Menken employed a stratified random sample technique to select the 10 New York high schools in the study sample. The author described an unexpected effect of test washback[8] from her interviews with bilingual teachers and her observations of Spanish language classes at one high school. Menken reported that educators from the New York City high school realized that skills on the Spanish

[8] Test washback is defined as the impact of a high-stakes test on classroom instruction (Wall & Horák, 2007).

Advanced Placement (AP) curriculum and exam were the same skills required to pass the English Regents exam. Thus, the Spanish AP course became a requirement and test preparation strategy for Hispanic ELLs taking the English Regents exam.

Limitations

The present study is limited in four areas. First, inclusion criteria limited research studies to those studies published in peer-reviewed journals. Thus, the sample of reviewed studies may not be as representative of the corpus of research literature on high-stakes testing and ELLs, as it might have been if other published sources of research studies, (e.g., government reports, dissertation studies) were included in the sample. Secondly, the search protocols and procedures followed may not have captured all relevant studies because studies were not indexed in the databases searched or cited in relevant articles. Thirdly, the MQS instrument may favor some research designs (i.e., longitudinal over cross-sectional) and analytical approaches over others (i.e., multivariate over univariate). This bias may disadvantage research paradigms that do not generally employ these techniques (i.e., qualitative paradigms) or advantage others (i.e., mixed methods). Finally, the MQS is a qualitative instrument that represents my bias with regard to the criteria selected for evaluating each study. The MQS has not been previously validated. However, it should also be noted that these criteria were adapted from other MQS instruments.

Implications for Practice

ELLs' English literacy acquisition is a major issue for all public education stakeholders. Consider the following demographics. First, one in five U.S. school

children is the son or daughter of immigrants. The majority of these second generation Americans are Hispanics whose primary language is Spanish. Second, 21% of second-generation Hispanic children (i.e., U.S.-Born children of one or both immigrant parents) are not fluent in English. For first-generation Hispanic children (i.e., non-U.S.-Born children), the number increases to 43% (Fry & Passel, 2009). Third, in the year 2000, ELLs represented 6% of U.S. school-age children (Capps et al., 2005). By the 2003-2004 school year, 11% of U.S. school-age children received English language learner services (NCES, 2006). In this section, I will discuss implications for practice based on the current state of the literature in the area of ELLs' English literacy acquisition and high-stakes testing (i.e., NCLB's accountability by assessment).

The present review identified only11 eligible studies. Thus, these studies function as the corpus of evidence for the current state of the literature about high-stakes testing and ELLs. Moreover, this evidence presents a fragmented picture of ELLs' academic achievement, performance, and outcomes. First, the small sample size (n=11) points to a paucity of research literature about the following: (a) the assessment of ELLs' academic achievement utilizing standardized tests normed on English-speaking students, (b) the impact of assessment outcomes on ELLs (i.e., retention policies, graduation policies, *de facto* language policies created by LOI allotments in classrooms), (c) and English language acquisition trajectories. Secondly, while a myriad of questions are suggested by the research evidence from the present study, two overarching questions should be considered by education policymakers and other education stakeholders in the reauthorization of NCLB and in deciding the U.S. government's future role in

assessment and external accountability. Question 1: *At what point in ELLs' English language acquisition trajectory, is their performance on high-stakes assessments, normed on English-speaking students, a measure of their content-area literacy and not a measure of their English language proficiency?* A national assessment normed on English-speaking and ELL students might provide a more accurate picture of academic achievement and gains for EOs and ELLs. Question 2: *Should high-school exit assessments in the primary language of students (i.e., non-English languages) be allowed in addition to the required English assessment?* An example of content area assessments in a language other than English is the New York Regents exams. Currently, there are several states that offer state-wide assessments in a language other than English in primary but not secondary grades. The issue at the heart of both questions is the validity of ELLs' test results.

Noticeably absent from the literature on high-stakes testing are dual language or two-way immersion programs. Generally, in dual language programs grade-level cohorts consist of EOs or English fluent students and ELLs. These programs are considered additive, meaning that students become bilingual and biliterate (i.e., develop native or near native proficiency in both languages). Empirical studies conducted with ELLs and EOs from these programs should examine crosslinguistic transfer for both EOs and ELLs and their academic performance on high-states assessments.

As noted previously, the present review contained only two studies of secondary ELLs, one from middle school and one from high school. Academic and content area literacy development and their assessment are especially important for secondary ELLs.

The depth and breadth of domain-specific language and text structures required to pass high-stakes tests increase exponentially in secondary grades. Similarly, no study addressed linguistic accommodations (i.e., the use of language dictionaries, reading test items to students), another critical assessment issue for secondary ELLs. Moreover, questions remain regarding the impact of linguistic accommodations on the validity of test results.

Finally, given the availability of student data warehoused by state agencies, educational researchers should consider the analysis of secondary data. Access to student data, including student demographic information and academic performance data, would allow researchers to employ statistical analysis approaches, such as multilevel techniques, that require large sample sizes. Additionally, accessibility and training seminars in the use of the states' databases could be an effective data management tool for state agencies.

CHAPTER III

ENGLISH ORAL LANGUAGE DEVELOPMENT AND ENGLISH LANGUAGE LEARNERS: A SYSTEMATIC REVIEW OF THE RESEARCH EVIDENCE

Introduction

Beginning in 1996, a major synthesis of the literature on ELLs' language and academic development was conducted by research teams from the Center for Research on Education, Diversity and Excellence (CREDE). Research was organized around six topics: language learning and academic achievement; professional development; family, peers, and community; instruction in context; integrated school reform; and assessments. The results were published in *Educating English Language Learners: A Synthesis of Research Evidence* (Genesse, Lindholm-Leary, Saunders, and Christian 2006). My study was a follow-up study of one of the subtopics of the Genesse et al. study (2006). In my study, I reviewed empirical studies on oral language development and ELLs, one of the subtopics under language learning and academic achievement from the period not included in Genesse et al. (2006), 2003 through 2009 (May). In the next section, I will provide a brief summary of the Genesse et al. (2006) synthesis.

The following inclusion criteria defined the search of the literature for the reviewed studies in Genesse et al. (2006): (a) empirical; (b) conducted in the United States; (c) published in English; (d) examined oral language development, literacy, and academic achievement among ELLs with outcome measures in English; (e) centered on pre-kindergarten through twelfth grade; (f) published during the past 20 years with

earlier seminal studies included; (g) published in peer-reviewed journals with some technical reports included; and (h) for literacy studies, contained literacy related outcomes (i.e., reading and writing). The synthesis comprised four domains: oral language, crosslinguistic and crossmodal issues in literacy, instructional issues in literacy, and achievement. In addition to its domain designation, oral language was also a factor in two other domains, crosslinguistic and crossmodal issues in literacy and achievement. A total of 182 studies from the years 1980-2003 were reviewed.

Evidence from the Genesse et al. (2006) study regarding oral language development and its association to literacy and achievement suggested that oral language incorporated into academic language correlated with higher levels of oral language proficiency and literacy achievement. Secondly, primary language proficiency (L1) did not detract from and appeared to support developing literacy in L2. This conclusion was supported by findings from the best evidence synthesis of language of reading instruction for ELLs conducted by Slavin and Cheung (2005). Of the 13 studies reviewed by Slavin and Cheung (2005), 9 favored bilingual approaches, whereas no study results favored English-only (i.e., immersion) approaches. Third, evidence from Genesse et al. (2006) suggested that the associative relationship between academic oral language and reading achievement appeared to grow stronger in advancing grades. In conclusion, based on their synthesis of the literature, Genesse et al. pointed to the lack of research evidence on English oral language development of ELLs. Furthermore, the authors reported the need for evidence-based comprehensive frameworks that would

provide a coherent plan for explicit instruction of specific oral language skills and subskills in ELL classrooms.

My study contributes to research evidence on the topic of oral language and ELLs in three ways. First, to my knowledge no systematic, replicable review of the literature on oral language and ELLs has been conducted since the publication of the Genesse, Lindholm-Leary, Saunders, and Christian study (2006). Second, there are no studies to date that have evaluated the methodological quality of studies focusing on oral language and ELLs. Third, the studies conducted by Genesse et al. (2006) contributed to our knowledge about ELLs' language and achievement by its comprehensive nature (i.e., published in the past 20 years). However, the conclusions and findings of Genesse et al. are based on studies of language development and ELLs (n=48) that are dated. Over 87% of the studies were *published* before 2000. Thus, the studies in my review offer a more current state of the literature on oral language and ELLs.

Purpose

My study examines current research evidence on oral language development and ELLs and their progress towards academic achievement parity with their native-English speaking peers. The goal of my study is to evaluate the validity and generalizability (quantitative research) or trustworthiness and transferability (qualitative research) of the research findings. A secondary goal is to categorize the themes and issues derived from the research questions of the abstracted studies.

The literature on oral language and ELLs was systematically reviewed to answer this question.

1. What are the methodological characteristics of the literature on oral language and ELLs published in the years 2004-2009?

Conceptual Framework

Snow defines oral language as "all oral forms of communication, speaking, and listening" (p. 166). For ELLs, as for all students, three aspects of oral language are crucial to academic achievement: (a) as social interaction, (b) as cognitive tool for participating one's own learning, and (c) as portal to academic discourse (Saville-Troike, 1989).

Snow (1991) initially proposed a conceptual model of literacy development in which decontextualized language skills, developed from kindergarten through second grade, would predict student reading performance through fourth grade for EO students. Proctor, Carlo, August, and Snow (2005) hypothesized a similar conceptual model for ELLs. They found that for fourth grade ELLs, English oral language (L2) skills were better predictors of reading achievement outcomes than decoding skills. Studies reviewed in my study generally operationalized oral language as verbal knowledge or as verbal knowledge and listening comprehension.

Method

This systematic review of the literature consisted of two parallel systematic reviews of empirical studies. The first systematic review examines ELLs' performance and its relationship to high-stakes assessments was presented in Chapter II. The second systematic review, a follow-up study of an earlier research literature synthesis conducted by Genesse et al. (2006), examined ELLs' oral language and the relationship to literacy.

My review was conducted using a modified version of Garrard's (2007) Matrix Method for the health sciences and Slavin's (1986, 1995) best evidence approach for educational scientific research syntheses. Three sources were employed to locate scientific journal articles: electronic databases, purling, and journal searches. Once empirical studies addressing these relationships' were selected, articles were abstracted and analyzed using the Garrard matrix and coding sheets constructed for this systematic review. Modifications to the coding sheets in my study were informed by Honoré's (2008) systematic review of genetic risk and mate selection research literature in the health sciences. Research findings were synthesized and categorized according to the research questions found in the studies. These synthesized findings formed the basis for the present systematic review.

Inclusion and Exclusion Criteria

The empirical studies considered for inclusion in the systematic review of research evidence on oral language and ELLs were selected based on seven inclusion criteria. The articles had to: (a) be empirical studies (qualitative or quantitative); (b) be conducted in the United States; (c) be published in English; (d) be centered on some measurement of ELLs' oral language and literacy; (e) involve ELLs from prekindergarten through twelfth grades; (f) be published in the period from 2004 through 2009; and (g) be published in a referred journal. The current systematic review of oral language and ELLs is a follow-up study of Genesse et al. (2006) synthesis of evidence-based research on oral language and ELLs. This earlier review included articles

published through 2003. Therefore, studies eligible for inclusion in the present study were published after 2003.

Studies were excluded from the oral language review if they: (a) were reviews of the research literature, book reviews, commentaries, editorial articles, policy analyses, theoretical or framework proposals, unpublished manuscripts, and letters to the editor; (b) focused only on oral language development and did not explicitly report a relationship between oral language development and literacy; (c) focused only phonological awareness, print awareness, and/or elements of grammar (e.g. syntax, morphology) without including oral language development and literacy; (d) focused only on language proficiency testing without including oral language development and literacy; (e) did not explicitly include ELLs in the sample and report finding for ELLs; and (f) were case studies of fewer than three ELLs.

Search Techniques

Four electronic databases were employed for article search and retrieval. These databases included ERIC (CSA), Academic Search Complete (Ebsco), PsycInfo (CSA), and Education Full Text (Wilson). Search strategies for oral language and ELLs employed variations of Boolean key terms and database descriptors. Search words were: *English language learners* and *oral language*: *second language learners, English second language, (second or English) and language, second language learning, language minority, expressive language, language acquisition.* (See Figure 1, Chapter II.)

Articles were screened and retrieved based on their titles and abstracts. Qualifying studies missed during the Boolean term search were supplemented by purling

the reference page of articles found from the searches and were added to the extant corpus of research studies. Articles gleaned from purling were also located via the searches of the electronic databases and the manual search of the journals (see Appendix D). Of the 202 studies retrieved and reviewed, the total sample of qualifying studies for oral language and ELLs from the electronic database search and the manual search of the journals was 23 articles (*N*= 23; see Appendix C-2 for a list of abstracted studies). All retrieved studies were assigned a number and inventoried in an Excel spreadsheet.

Data Abstraction and Analysis

The corpus of empirical studies included in the present systematic review of the research literature was generated in four steps: (a) a search of electronic databases in December 2008; (b) an examination of purled reference sections of retrieved articles; and (c) a manual article search by individual journal's table of contents; and (d) a second search of electronic databases in May 2009 to expand and locate any articles published between December 2008 and May 2009. In addition, the second search of electronic databases replicated the first search by combining the high-stakes and oral language searches (See Figure 3).

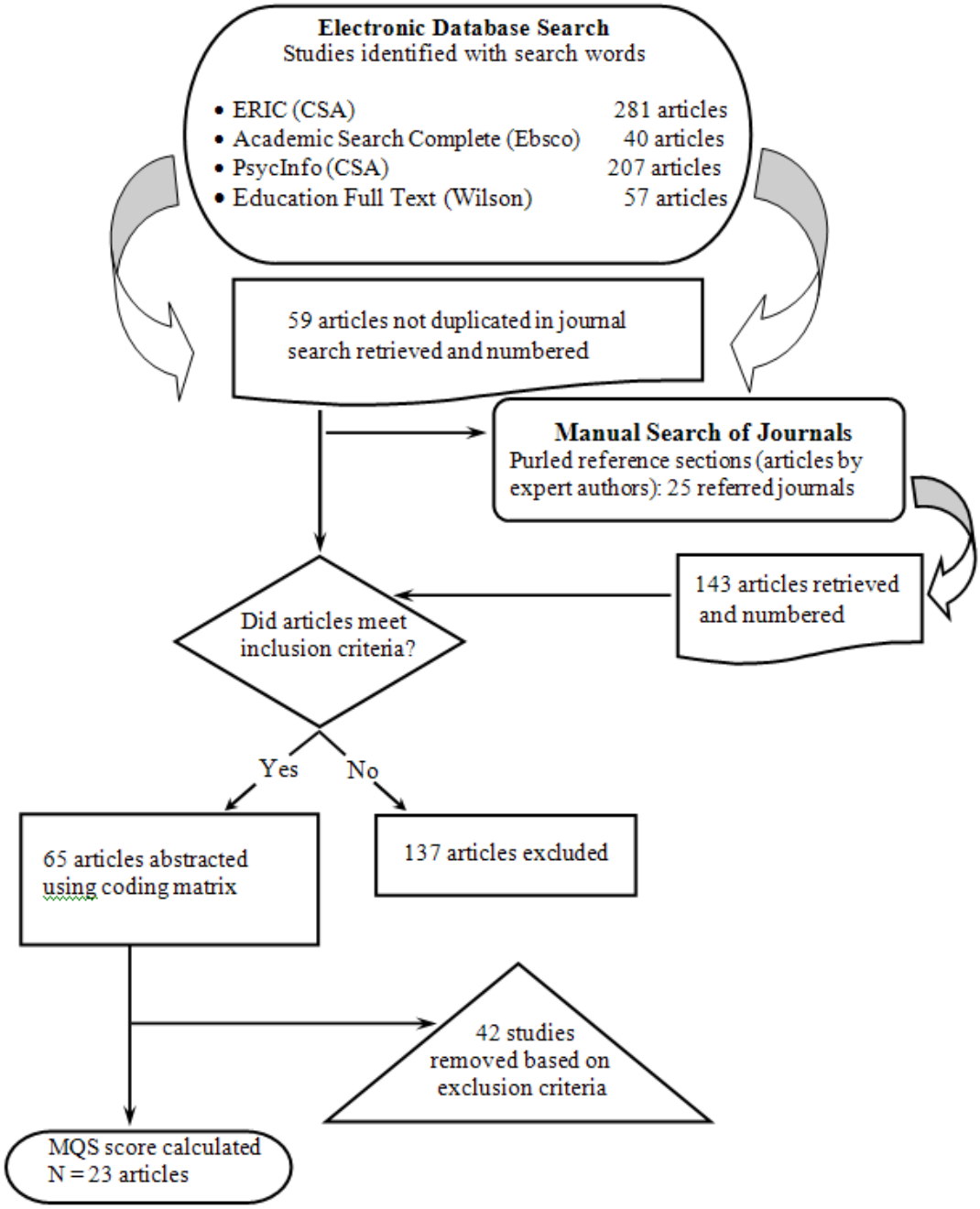

Figure 3. Methodological Steps: Oral Language and ELLs Systematic Review

The peer-reviewed journals for the manual search were selected from articles published in the years 2004 through 2009 by experts in the field (e.g., August, Snow). From purled journals, 25 journals that most frequently appeared in the reference sections were selected for this journal search. The table of contents of all issues published in the years 2004 through 2009 was inspected for each selected journal. One hundred thirty-seven articles were retrieved. Of those articles, 17 studies met the selection criteria in the final sample (*N*=23).

Methodological Quality Assessment

The MQS for the oral language and ELLs sample ranged from a minimum of 12 points to a maximum of 19 points. Table 3 presents the MSQ criteria with a frequency distribution of the individual studies for each criterion scoring category. Each reviewed article's individual MQS criterion score and its total MQS is listed in the summary contained in Table 4. (For a detailed description of the methodological assessment see Chapter II. Oral language and ELLs sample abstraction sheet with MQS scoring rubric are presented Appendix E-2.)

Table 3.

Summary of Frequency of MQS Criteria Scores for Oral Language and ELLs Abstracted Studies (N=23)

Criteria	Mixed Methods Studies %	Quantitative Studies %	Total MQS Score Frequency %
C1- Operationalization of variables:			
No	50.0	33.3	**34.8**
Partial	00.0	14.3	**13.0**
Yes	50.0	**52.4**	***52.2***
C2- Theory:			
No	50.0	**52.4**	**52.2**
Yes	50.0	47.6	***47.8***
C3-Instrumentation:			
Reliability and validity unreported	50.0	9.5	**13.0**
Reliability and validity reported	50.0	**90.5**	***87.0***
C4-Research paradigm:			
Qualitative			**00.00**
Quantitative			***91.3***
Mixed Methods			**8.7**
C5-Research design:			
Idiographic/qualitative	00.0	00.0	**00.0**
Cross-Sectional	50.0	14.3	**17.4**
Cohort: Pre-Posttest	50.0	61.9	***60.9***
Cohort: ≥3 waves	00.0	23.8	**21.7**
C6-Sampling:			
Non-random	**100.0**	**57.1**	***60.9***
Non-random and comparison group	00.0	00.0	**00.0**
Random	00.0	42.9	**39.1**

Table 3. *(continued)*

Criteria	Mixed Methods Studies %	Quantitative Studies %	Total MQS Score Frequency %
C7-Sample size:			
Small - qual. = <30	00.0	00.0	**00.0**
Small - quan.= <30	00.0	00.0	**00.0**
Medium = >30 <300	50.0	**61.9**	***60.9***
Large = >300	50.0	38.1	**39.1**
C8-Language groups:			
ELLs	50.0	**71.4**	***65.2***
ELLs and EOs	50.0	9.5	**13.0**
ELLs-Dual Language Programs[a]	00.0	19.0	**21.7**
C9-Participants:			
Early childhood, P-2	50.0	**66.7**	***65.2***
Upper elementary, 4-6	50.0	9.5	**13.0**
Primary school, P-6	00.0	23.8	**21.7**
Middle school, 6-8	00.0	00.0	**00.0**
High school, 9-12	00.0	00.0	**00.0**
C10-Data analysis:			
Univariate	50.0	4.8	**8.7**
Bivariate	00.0	00.0	**00.0**
Multiple regression	00.0	57.1	***52.2***
Multivariate statistics (i.e., HLM, SEM)	50.0	38.1	**39.1**

Note. See Appendix E-2 for abstraction sheet descriptors. Figures may not add to 100% due to rounding.
[a]ELLs in dual language (DL) program models were included in the sample. DL models typically have EO (or fluent English speakers) and ELLs in the same cohort.
EO = English Only, fluent English-speaking, i.e., non-English language learner, who may or may speak or be fluent in other non-English languages; qual. = qualitative study; quan. = quantitative study.

Table 4.

Evaluation Matrix: MQS Review Criteria Scores of Oral Language and ELLs Abstracted Studies

Reference		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	MQS total score
Mixed methods studies (MM):												
López & Tashakkori (2006)	MM	2	0	2	2	1	1	2	1	2	4	17
Spycher (2009)	MM	0	2	1	2	2	1	1	1	1	1	12
Total mixed methods studies	**M**	**1.00**	**1.00**	**1.50**	**2.00**	**1.50**	**1.00**	**1.50**	**1.00**	**1.50**	**2.50**	**14.50**
Range (12-17), Median=14.5	**SD**	**1.00**	**1.00**	**0.50**	**0.00**	**0.50**	**0.00**	**0.50**	**0.00**	**0.50**	**1.50**	**2.50**
Quantitative studies (Qn):												
Cárdenas-Hagan et al. (2007)	Qn	0	2	2	1	2	3	2	2	1	3	18
Cirino et al. (2007)	Qn	1	0	2	1	2	1	2	2	1	3	15
Gottardo et al. (2008)	Qn	2	0	2	1	2	1	1	1	1	3	14
Gottardo & Mueller (2009)	Qn	2	2	2	1	2	1	1	1	1	4	17
Kelly et al. (2008)	Qn	0	0	1	1	2	1	2	1	1	4	13
Kieffer (2008)	Qn	2	2	2	1	3	1	2	1	1	4	19
Laija-Rodríquez et al. (2006)	Qn	2	2	2	1	3	1	1	1	1	3	17
López & Tashakkori (2006)	Qn	2	0	2	2	1	1	2	2	2	4	18
Manis et al. (2004)	Qn	1	0	2	1	3	1	2	1	1	3	15
Miller et al. (2006)	Qn	2	0	2	1	1	1	2	1	1	3	14
Nakamoto et al. (2007)	Qn	1	0	2	1	3	3	1	1	1	4	17

Table 4. *(continued)*

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	MQS total score
Nakamoto et al. (2008)	Qn	0	0	1	1	2	3	1	1	1	4	14
Proctor et al. (2005)	Qn	2	2	2	1	1	1	1	1	2	4	17
Proctor et al. (2006)	Qn	2	2	2	1	1	1	1	1	2	3	16
Roberts (2008)	Qn	0	2	2	1	2	3	1	1	1	3	16
Saunders et al. (2006)	Qn	2	0	2	1	2	1	2	2	1	3	16
Silverman (2007)	Qn	0	0	2	1	2	1	1	2	1	4	14
Tong et al. (2008a)	Qn	2	0	2	1	2	3	1	1	1	1	14
Tong et al. (2008b)	Qn	0	0	2	1	3	3	2	1	1	4	17
Vaughn et al. (2006a)	Qn	2	0	2	1	2	3	1	1	1	3	16
Vaughn et al. (2006b)	Qn	0	2	2	1	2	3	1	1	1	3	16
Vaughn et al. (2006c)	Qn	0	2	2	1	2	3	1	1	1	3	16
Total quantitative studies	**M**	**1.14**	**0.82**	**1.91**	**1.05**	**2.05**	**1.82**	**1.41**	**1.23**	**1.14**	**3.32**	**15.86**
Range (13-19), Median=16	**SD**	**0.92**	**0.98**	**0.29**	**0.21**	**0.64**	**0.98**	**0.49**	**0.42**	**0.34**	**0.70**	**1.55**
Total abstracted studies	**M**	**1.13**	**0.83**	**1.88**	**1.13**	**2.00**	**1.75**	**1.42**	**1.21**	**1.17**	**3.25**	**15.75**
Range (12-19), Median=16	**SD**	**0.95**	**1.01**	**0.34**	**0.34**	**0.66**	**0.99**	**0.50**	**0.41**	**0.38**	**0.85**	**1.73**

Note. See Appendix E-2 for abstraction sheet criterion descriptors.
Qn = Quantitative; MM = Mixed Methods. Median=16; mode= 16, 17; skewness= -0.30

Results

The search of empirical journal articles published in peer-reviewed journals between 2004 and 2009 was conducted in two phases (December 2008 and May 2009) and included a manual search of 25 peer-reviewed journals. This dual-phase research yielded 765 articles related to oral language and ELLs. Article abstracts were scanned resulting in 202 articles that were retrieved and numbered.

The final sample (*N*=23) was selected through a two-tier process. The first tier selection process, which yielded 65 articles, consisted of reviewing each article and afterwards completing a coding sheet of inclusion criteria. The second review to ensure that all articles met both the inclusion and exclusion criteria yielded 23 articles (see Appendix E-2 for abstraction sheets). Articles were excluded when they did not meet the inclusion criteria (e.g., were not empirical studies, did not relate directly to oral language and academic achievement for ELLs, focused on populations other than pre-kindergarten through twelfth grade students, or were conducted outside the U.S.).

Studies' Characteristics

Of the 23 reviewed studies, two-thirds of the articles were published in the years 2006 (34.8%) and 2008 (30.4%). The remaining studies were distributed over the years 2004 (*n*=1), 2005 (*n*=1), 2007 (*n*=4), and 2009 (*n*=2). Generally studies (65.2%) were part of larger research projects (Cárdenas-Hagan, Carlson, & Pollard-Durodola, 2007; Cirino, Pollard-Durodola, Foorman, Carlson, & Francis, 2007; Miller et al., 2006; Vaughn et al., 2006a; Vaughn et al., 2006b; Vaughn et al., 2006c).

Studies appeared in various professional journals and represented diverse disciplines and specializations in education: school psychology (*n*=4), special education (*n*=5), and reading (*n*=3). Other journals included *The Elementary School Journal* (*n*=5) and the *American Educational Research Journal* (*n*=2). Of professional journals specific to Hispanics or ELLs, two reports were found in *Bilingual Research Journal (BRJ)*, an official publication of the National Association for Bilingual Education (NABE). One report was found in *Hispanic Journal of Behavioral Sciences*; and one was found in *TESOL Quarterly,* published by Teachers of English to Speakers of Other Languages, Inc. (TESOL).

Generally, reviewed studies presented literacy theories implicitly (52.2%, *n*=12) rather than establishing theoretical links to constructs by means of a theoretical framework. Over one third of the studies (39.1%, *n*=9) examined crosslinguistic transfer of L1 skills to L2 as a factor in ELLs' literacy acquisition. Thus, the predominant theories cited in studies with implied and explicit theoretical linkages to the crosslinguistic transfer construct were Cummins' *Developmental Interdependence Hypothesis* (Cárdenas-Hagan et al., 2007; Gottardo & Mueller, 2009), *Central Hypothesis (Common Underlying Proficiencies)* (Cárdenas-Hagan et al., 2007), and *Threshold Hypothesis* (Laija-Rodríguez, Ochoa, & Parker, 2006) theories. The second most cited theory was the *Simple View of Reading* (Gottardo & Mueller, 2009; Proctor, August, Carlo, & Snow, 2006), a conceptual framework of factors that predict reading performance. Other theories presented were *Differential Skills Hypothesis* and *Developmental Lag Hypothesis* (Kieffer, 2008), *Central Processing Hypothesis*

(Gottardo & Mueller, 2009), *Foundational Theory* (Proctor et al., 2006), and *Systemic Functional Linguistics* (Spycher, 2009).

Studies' Methodological Quality

Each study from the oral language and ELL sample (*N*=23) was evaluated and assigned a methodological quality score (MQS), which was a composite of the 10 criteria scores (for an example of the MQS protocol, see Appendix E-2). Table 3 presents the frequency distribution summary of the MQS criteria. MQS scores for reviewed studies ranged from 12 points to 19 points for the current sample (Table 4). Additionally, studies were grouped by research paradigm (i.e., quantitative, mixed methods) with MQS scores reported as subtotals. The mean total MQS for all reviewed scores was 15.75 (SD=1.73). Variability was greater in the mixed methods studies (*n*=2) than the quantitative studies (*n*=21) as indicated by the standard deviation of the mixed methods studies (see Table 4). The mean, median (median=16), and mode (mode=16, 17) total values were almost identical with a slightly negative normal distribution (skewness=-.30) of the scores. Studies' MQS was somewhat above (39.1%; *n*=9) the theoretical midpoint of 16 points with 34.8% (*n*=8) below and 26.1% (*n*=6) at the theoretical midpoint of 16 points.

Most of the reviewed studies employed a quantitative approach (91.3%, *n*=21). While no studies were qualitative, two studies used a mixed methods paradigm (*n*=2). Of these two studies, only one self-reported as mixed methods (López & Tashakkori, 2006). The second study (Spycher, 2009) was labeled mixed methods because of companion qualitative and quantitative studies or phases. The research designs employed in the

mixed methods studies were pretest-posttest (*n*=1) and cross-sectional (*n*=1). Generally quantitative studies employed pretest-posttest research designs (60.9%, *n*=14) with the remaining studies using either longitudinal (21.7%, *n*=5) or cross-sectional designs (17.4%, *n*=4).

More than one half (60.9%) of the sample sizes were medium (over 30 and under 300 participants) with no study having fewer than 30 participants. Studies predominantly used non-randomized sampling (60.9%, *n*=14). Additionally, kindergarten students were the most commonly represented group as a single grade level (21.7%) followed by cohort groups (39.1%) that ranged from kindergarten through first grade to kindergarten through sixth grade. While there was some variation in the primary school grade levels, no studies included samples from middle school or high school. Most studies (60.9%) did not report data collection years.

Overall, non-English primary language groups were predominantly Spanish-speaking ELLs (82.6%, *n*=19). Other primary language groups represented were Hmong, East Asian[9], and Haitian Creole. Studies with ELL-only samples represented 69.6% of the reviewed studies (*n*=16), and three studies (13.0%) compared ELLs and EOs (i.e., fluent English speakers). Four reports included participants from dual language classes in their sample (Cárdenas-Hagan et al., 2007; Cirino et al., 2007; Saunders, Foorman, & Carlson, 2006; Silverman, 2007).

Data collection techniques generally consisted of standardized tests (87%, *n*=20). Other data collection instruments were interviews (*n*=2), observations (*n*=3), and

[9] East Asian was a generic language term used by the authors.

questionnaires (*n*=1). Three studies utilized researcher developed measures. Spycher (2007) employed researcher developed science assessments, Emergent Science Vocabulary (ESVA) and Conceptual Interview on Scientific Understanding (CISU) in a pretest-posttest design conducted with ELL and EO kindergarten students. Likewise, Roberts (2008) developed vocabulary assessments for a storybook reading home literacy intervention with Spanish- and Hmong-speaking ELL preschool students; and Silverman (2007) developed a vocabulary assessment for comparing vocabulary acquisition rates of ELL and EO kindergarten students. Two of the studies reported reliability scores for the researcher development assessments (Roberts, 2008; Silverman, 2007). Although Spycher (2007) reported that reliability statistics had been calculated, no reliability statistics were found in the study. Ninety percent (*n*=19) of the qualitative studies reported validity and reliability statistics, whereas only one mixed method study reported validity and reliability statistics. In contrast, slightly more than one half (*n*=13) of the reviewed studies operationalized variables or phenomena. However, studies that provided operational definitions did not employ the word operationalize or its derivative.

Statistical techniques for data analysis tended to be multiple or logistic regression (47.8%, *n*=11) followed by studies reporting specific analytical approaches (39.1%, *n*=9). Reported data analysis approaches were as follows: structural equation modeling (*n*=3), linear growth modeling (*n*=3); hierarchical level modeling (*n*=1), and multivariate analysis of variance (*n*=2). The remaining quantitative studies employed univariate (*n*=3) analyses.

Studies' Empirical Findings

The reviewed studies contained a total of 276 findings with a mean of 12.0 findings per study. ELLs' academic achievement was a dependent variable in two studies (Kieffer, 2008; López & Tashakkori, 2006). Phonemic awareness, phonetic skills, and oral language were the variables (predictor and outcome) measured in the majority of studies (53.8%, *n*=13). The phenomena examined in these studies were crosslinguistic transfer (46.2%, *n*=6), reading acquisition for struggling ELL readers (23.1%, *n*=3), second language acquisition models (15.4%, *n*=2), and reading models for ELLs (15.4%, *n*=2). Other phenomena examined were teacher quality (*n*=1) and English as a secondary instruction allotments within the language arts time block (*n*=1).

A preliminary review of the findings identified two major trends in EL's English (EL2) literacy development. The first trend was ELL's primary language skills as predictors of EL2 literacy development. Generally samples from these studies were drawn from Spanish-speaking ELLs with one sample including Spanish- and Hmong-speaking ELL preschool children (Roberts, 2008). Overall, studies (*n*=8) found some level of cross-language association between English and the primary language. These associations appeared strongest in early primary bilingual classrooms where Spanish was the language of reading instruction (Cárdenas-Hagan et al., 2007; Laija-Rodríquez et al., 2006; Manis, Lindsey, & Bailey, 2004; Miller et al., 2006). The exception was Robert's study (2004) of home literacy practices in a preschool English immersion (EI) classroom. Evidence supported the use of primary language storybooks to increase ELL's L2 targeted vocabulary.

The second trend was the strength of association between L2 oral language and L2 reading comprehension in reading literacy models for ELLs. While findings supported the relationship between oral language and reading comprehension across grade levels (Gottardo & Mueller, 2009; Miller et al., 2006), generally reading decoding and fluency skills had a stronger association to reading in the early primary grades (Cárdenas-Hagan et al., 2007). For studies conducted in the upper elementary grades, findings diverged. Employing the *simple view of reading* model, Proctor et al. (2006) conducted a non-experimental study of ELL fourth graders. Findings showed a stronger relationship between L2 oral language (i.e., listening comprehension) and L2 reading comprehension than between L2 decoding skills and L2 reading comprehension. Conversely, third grade decoding skills continued to be stronger predictors of reading comprehension than oral language skills for Spanish-speaking ELL 6th graders in both Spanish and English (Nakamoto, Lindsey, and Manis, 2008).

Discussion

This review contributes to research evidence on oral language and ELLs in two ways. First, the review focuses on research published in 2004-2009, which is the period following the publication of a major research synthesis on educating ELLs (Genesse et al., 2006). Secondly, this systematic review of the literature, focusing on the methodological quality of the corpus of studies, permits a replicable, critical evaluation of the literature that is absent in nonsystematic or narrative reviews.

A secondary aim of my study was to examine themes resulting from patterns identified in reviewed studies' research questions. Study findings were categorized by

the research question. From these categories emerge two trends. The first trend in the findings was studies that examined cross-language skill transference. The second trend was studies that examined EL2 reading comprehension performance predictors.

Crosslinguistic Transfer

Overall studies found some level of cross-language transfer of L1 to L2 reading skills. Cárdenas-Hagan et al. (2007) examined the relationship between L1reading skills and the same skills in L2 in kindergarten ELLs assigned to English immersion, dual language, and transitional bilingual education classrooms. Evidence supported the L1 transference of phonemic awareness and phonetic skills across languages in ELLs with lower levels of L2 proficiency but not in ELLs with higher levels of L2 proficiency. Manis et al. (2004) examined crosslinguistic transference from first grade to second grade in a cohort of Spanish-speaking ELLs. The authors found that L1 expressive language (i.e. word identification), print, and phonemic awareness skills predicted L2 reading passage scores. Additionally they also found evidence of L2 to L1 transference. Secondary language phonemic awareness skills in first grade corresponded to second grade L1 reading passage scores. In a second study Miller et al. (2006) examined crosslinguistic transference of oral language skills in a cross-sectional study of K-2 grade ELLs. They found that L1 oral language skills predicted L2 passage comprehension scores. Likewise, L2 oral language predicted L1 passage comprehension. However, the associations were weak with R2 indexes reported as 2% for L1 oral language to L2 passage comprehension and 6% for L2 oral language to L1 passage comprehension.

Nakamoto, Lindsey, and Manis (2008) examined cross language transfer from third grade to sixth grade of Spanish-speaking ELLs in bilingual transition education program models that were early exit. The authors noted that while the cross-language transfer of L1 to L2 reading comprehension was negligible, data model fit was improved when L1 reading skills were added as factors in the model. Likewise, Proctor et al. (2006) found that Spanish vocabulary knowledge enhanced L2 reading comprehension in fourth grade ELLs. The two studies (Nakamoto et al., 2008; Proctor et al., 2006) differed in their views of L1 and L2 alphabetic knowledge and fluency as representing a common skill. Proctor et al. (2006) proposed the common skill hypothesis for L1 and L2 alphabetic knowledge and fluency. Conversely in the Nakamoto et al. study (2008) and in a study conducted by Gottardo and Mueller (2009), model fit was improved when L1reading skills and analogous L2 reading skills were treated as separate factors.

EL2 Reading Comprehension Performance Predictors

While reviewed studies supported oral language as a predictor in reading comprehension models (Proctor et al., 2005; Miller et al., 2006; Nakamoto et al., 2008) differences in the strength of the association between oral language and reading comprehension performance were found. Generally findings supported Snow's (1991) model of the importance of decoding skills as reading performance predictors through 2nd grade. In the fourth grade, decoding skills would diminish in importance and oral language would increase in importance as predictors of reading comprehension performance (Cárdenas-Hagan et al., 2007; Gottardo & Mueller, 2009; Proctor, 2005). It should be noted that the Proctor et al. study (2005) is the only study of fourth grade

ELLs. The other two study participants ranged from kindergarten (Cárdenas-Hagan et al., 2007) through 1st and 2nd grades (Gottardo and Mueller, 2009). In a study of fifth grade Spanish-speaking ELLs, Nakamoto et al. (2008) found that third grade decoding skills continued to be stronger predictors of 6th grade reading comprehension performance than third grade oral language skills in both Spanish and English.

Limitations

For a detailed discussion of the limitation of my study see Chapter II.

Implications for Practice

The present review found only 23 eligible studies. Thus, these studies function as the corpus of evidence for the current state of the literature about oral language and ELLs. Two major trends in the findings on oral language and EL2 literacy were identified. The first trend was the transference of L1 oral language skills to L2 literacy skills. Evidence from this study supports the transference of literacy skills in the primary language to L2 reading. Additionally, these associations between L1 and L2 crosslinguistic transference appear to be the strongest in the early primary grades and where the primary language (e.g. Spanish) is the language of reading instruction. The second trend in the findings was oral language as a predictor of reading comprehension in EL2 literacy models. While evidence has supported oral language as a predictor of L2 reading comprehension, decoding skills appear to be the strongest predictors of reading comprehension in the early primary grades. Studies of ELLs in the upper grades present conflicting findings about the strength of association of these underlying skills.

Therefore, more studies are needed in the upper elementary and secondary grades to identify strong predictors of reading achievement older ELLs.

Noticeably absent from the literature on oral language and ELLs were qualitative studies and studies of secondary ELLs. Thus, two overarching questions should be considered for future research, including qualitative studies, on oral language and older ELLs (i.e., upper elementary and secondary grades). Question 1: *What are effective strategies to increase academic oral language in content areas?* Academic oral language development is important in two ways. First, academic oral language impacts ELLs' comprehension of domain-specific texts. Second, academic oral language development is crucial to classroom discourse about domain related concepts. Question 2: *What is the relationship of oral language component skills (i.e., vocabulary and listening comprehension) to EL2 literacy development in older ELLs?* Studies have operationalized vocabulary and listening comprehension skills as proxies for oral language. Findings support the association of vocabulary and listening comprehension to reading comprehension. While studies have examined vocabulary development in ELLs, no studies to my knowledge have examined listening comprehension development.

Finally, experimental studies of longitudinal data investigating EL2 literacy development have identified and examined important predictors of EL2 literacy development. However, as literacy competencies and standards change in advancing grade, ELLs' decoding skills will reach ceiling levels. Thus, identifying other predictors of literacy development (i.e. component skills of oral language) is essential to successful outcomes for ELLs.

CHAPTER IV

A LATENT GROWTH MODEL OF ENGLISH ORAL LANGUAGE TRAJECTORY AND HIGH STAKES READING ASSESSMENT FOR THIRD GRADE ENGLISH LANGUAGE LEARNERS

Introduction

Two major trends will determine educational policy and the role of government in educational reform for the next quarter century. The first is standards-based educational accountability that is data-driven and characterized by large-scale assessments. The second is the shift in the sociodemographic structures of U.S. public schools.

Assessment-based accountability is not a recent phenomenon in the United States. Indeed assessment of student performance has been a policy tool since the passage of the first Elementary and Secondary Education Act in 1965. (Popham, 2008) However, test-driven external accountability and proficiency testing, implemented under the No Child Left Behind Act of 2001 (NCLB), have become an important albeit controversial policy tool for educational reform (Lazear, 2006; Lee, 2008; Scheurich, Skrla, & Johnson, 2000; Valenzuela, 2005). State-wide large-scale assessments become *high-stakes* "when there are important contingencies linked to students' performances" (Popham, 2008; p. 2). For the individual test-takers, these contingencies can be grade-retention or failure to graduate. For teachers and administrators, they can be indicators of instructional quality (Popham, 2008).

One in five of school-age children are classified as bilingual or language minority, meaning that a language other than English is spoken at home. The majority of these students are Spanish-speaking (Fry & Passel, 2009). Thus, a major challenge for educators and schools is the persistent achievement gap that plagues Hispanic students, both English-speaking and English-learning, as one of the NCLB targeted groups. For this reason, identifying foundation skills that support literacy growth and improve student achievement outcomes is critical. Oral language with its array of components skills (i.e., vocabulary, listening comprehension) has been identified as one of those foundation skills (Proctor, Carlo, August, and Snow, 2005; Snow 1991).

Purpose

Secondary data from Project English Language and Literacy Acquisition (ELLA), a federally funded longitudinal study, was used to address the research questions in my study. The sample consisted of 185 Spanish-speaking ELL students enrolled continuously in the Project ELLA study from August 2004-May 2008 with no missing data. The data set consisted of scores from two Woodcock Language Proficiency Battery-Revised (WLPB-R) English Form (Woodcock, 1991a) subtests, Picture Vocabulary (PV) and Listening Comprehension (LC), and the third grade Texas Assessment of Knowledge and Skills in English (ETAKS) reading test. All students from the Project ELLA were assigned to four bilingual education instructional models: Structured English Immersion Enhanced/Experimental (SEI-E), Structured English Immersion Typical/Control (SEI-T), Transitional Bilingual Education Enhanced/Experimental (TBE-E), or Transitional Bilingual Education Typical/Control

(TBE-T). The present study examined the relationship between Hispanic ELLs' English oral language development over a 4-year period (kindergarten - third grade) and their outcomes on the state-mandated third grade ETAKS reading assessment. The goal was of my study was to determine whether the relationship between oral language and reading achievement identified in the research literature was also present when reading achievement was operationalized using a state-wide high-stakes reading assessment.

Research Questions

My study sought to answer these specific questions.

1. Does the English oral language proficiency trajectory of Hispanic ELLs explain their performance on a third grade state-wide high-stakes reading test in English?

2. Does ELLs' instructional program model assignment explain their English oral language trajectory?

Conceptual Model

My conceptual model was informed by reading literacy models for primary grades (for an overview of theories relevant to second language acquisition, see Appendix F). The first model, proposed by Snow (1991), was a reading comprehension performance model for English-speaking children. Parsing the underlying skills of student reading achievement, Snow (1991) hypothesized that the predictive strength of these skills would vary in advancing grades. Using information gleaned from research on emergent literacy and the development of literacy-based skills, Snow identified four skill domains developed during the preschool years (3-6 years) that would relate to literacy

development and reading achievement: conversational language, decontextualized oral language, print, and emergent literacy. This model captured the relationship between language and literacy development and their respective environmental supports, i.e., home and school. Snow hypothesized that pre-school print skills would predict student reading performance through second grade but would be less important predictors of reading performance in third and fourth grades. Conversely, while decontextualized language skills from kindergarten through second grade would be less important predictors of reading test performance in first and second grades, their predictive strength would increase in the higher grades.

In a second reading literacy model, Proctor, Carlo, August, and Snow (2005) conducted a cross-sectional study of 135 Hispanic ELLs in fourth grade. They examined the relationship of L2 oral language and decoding skills as predictors of L2 reading comprehension. Reading comprehension and oral language were operationalized using the following subtests of the Woodcock Language Proficiency Battery (WLPB): Passage Comprehension, Picture Vocabulary (PV), and Listening Comprehension (LC). Employing structured equation modeling (SEM) techniques to analyze their hypothesized model, EL2 reading comprehension was directly related to four factors: L2 listening comprehension, L2 vocabulary knowledge, L2 alphabetic knowledge, and L2 fluency. Additionally, verbal knowledge was indirectly related to reading comprehension through listening comprehension. The first two factors, listening comprehension and vocabulary knowledge, are oral language skills whereas alphabetic knowledge and fluency are decoding skills. While decoding played a minor predictive role in the fourth

grade Spanish-speaking ELLs' reading comprehension, EL2 oral language skills were statistically significant: listening comprehension (B=.44; $p<.001$); and vocabulary knowledge (B=.30; $p<.01$). The relationship between vocabulary knowledge and listening comprehension was also statistically significant (B=.85; $p<.001$). The authors reported an R^2 of .65. Thus, these findings appear to support Snow's (1991) hypothesis that print skills, i.e. decoding, and oral language skills present different associative patterns with reading performance at different developmental stages in the reading process.

The third piece needed to depict reading literacy development was a latent growth model fitted to longitudinal data, which could capture students' oral language growth trajectories. Tong, Lara-Alecio, Irby, Mathes, and Kwok (2008) hypothesized and tested a latent growth baseline model using three time points fitted to longitudinal data from Project ELLA. The Project ELLA study was a 4-year study of Hispanic ELLs from kindergarten through the completion of third grade. The authors examined the effectiveness of an English intervention in promoting EL2 academic oral language growth trajectory from kindergarten through the completion of first grade. Verbal knowledge and listening comprehension were proxies for oral language, which was operationalized utilizing the PV and LC subtests of the *WLPB-R*. Thus, results indicated that the model fit was adequate and that the EL2 oral language trajectory pattern over the two years was linear, one of the assumptions of latent growth models.

These baseline and conceptual models informed the conceptual framework for my conceptual model, which represented the associative relationship of EL2 oral

language to an English reading comprehension test. The secondary data in my study consisted of Project ELLA participants' scores from the PV and the LC subtests of the *WLPB-R* English Form (Woodcock, 1991a) and the ETAKS reading test. Figures 4 and 5 show the latent growth models for my study. The latent variable slope captured the oral language development trajectory. PV and LC scores were the predictor variables, and third grade ETAKS reading test scores were the outcome variable. PV (see Figure 4) and LC (see Figure 5) subtest scores were provided in five time points, the initial or baseline time point (time 0) taken in August 2004 of the participants' kindergarten year and four additional time points taken at 10-12 month intervals.

Method

Sampling and Research Design - Project ELLA

The present study used secondary longitudinal data from Project ELLA, a 6-year quasi-experimental study that was conducted in a large urban school district located in southeast Texas. Participants were young Hispanic ELLs, who were received instruction in one of four program models (SEI-E, SEI-T, TBE-E, and TBE-T) in kindergarten. It should be noted that schools rather than individual students were randomly designated as experimental or control. The control and experimental conditions continued through the participant's completion of third grade. Teachers and staff of Project ELLA collected data over a four year period from August 2004 through May 2008.

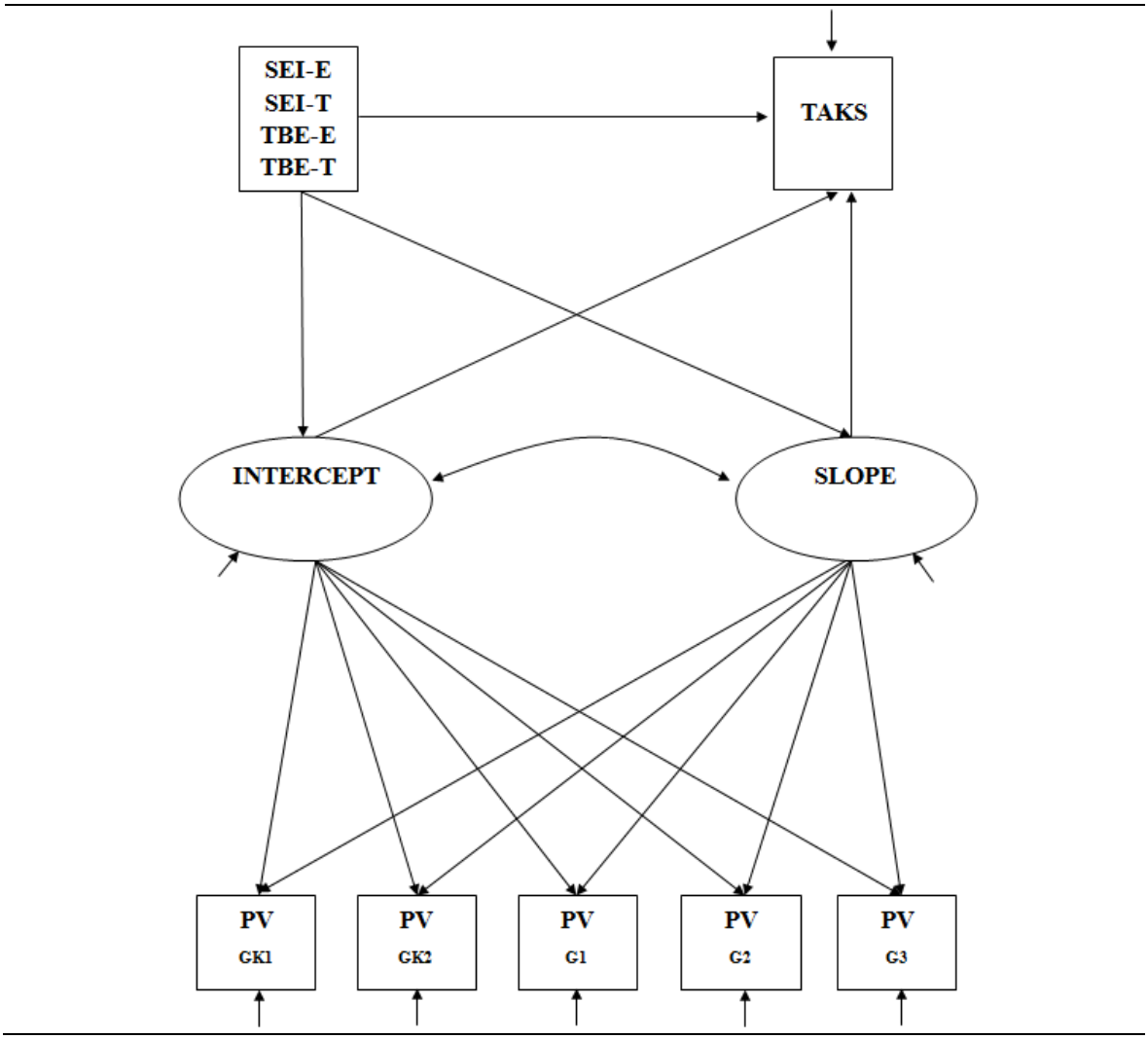

Figure 4. Latent Growth Model-Picture Vocabulary

Note. PV = Picture Vocabulary subtest of the Woodcock Language Proficiency Battery-Revised; GK1 = kindergarten, data collected in fall, 2004; GK2 = kindergarten, data collected in spring, 2005; G1 = 1st grade, data collected in spring, 2006; G2 = 2nd grade, data collected in spring, 2007; G3 = third grade, data collected in spring, 2008; SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control. TAKS = English TAKS reading test.

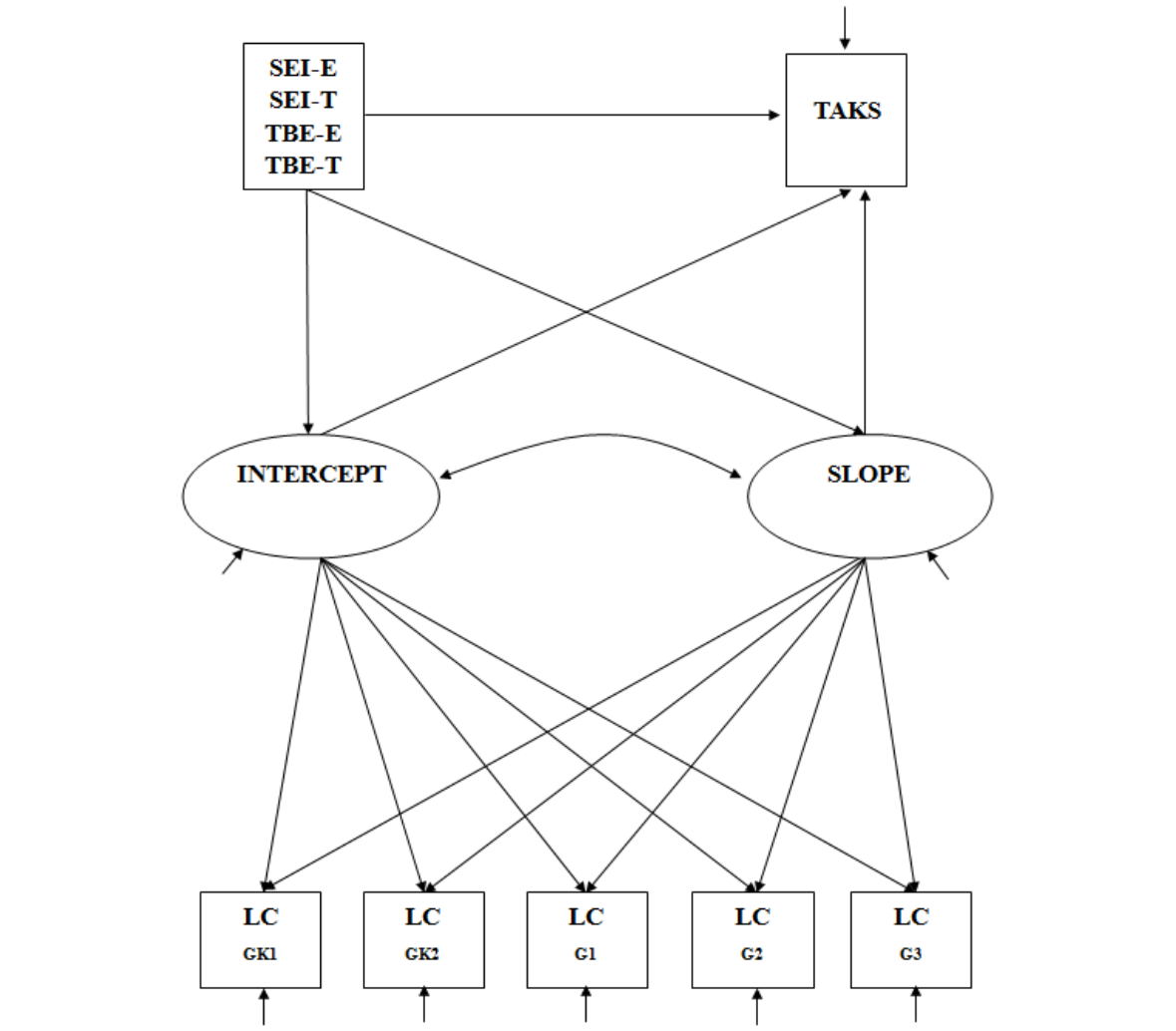

Figure 5. Latent Growth Model-Listening Comprehension

Note. LC = Listening Comprehension subtest of the Woodcock Language Proficiency Battery-Revised; GK1 = kindergarten, data collected in fall, 2004; GK2 = kindergarten, data collected in spring, 2005; G1 = 1st grade, data collected in spring, 2006; G2 = 2nd grade, data collected in spring, 2007; G3 = third grade, data collected in spring, 2008; SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control. TAKS = English TAKS reading test.

School Context

The setting for Project ELLA was an urban school district with the following demographics reported from the Public Education Information Management System (PEIMS) data for the school year 2007-2008. The two largest subpopulations were Hispanics (64%) and African American students (30%). White, Asian/Pacific Islander, and Native American accounted for approximately 5% of the student population. The majority of students (80%) were economically disadvantaged, and Limited English proficient (LEP) students represented one third of the student population. In the 2007-2008 Academic Excellence Indicator System (AEIS), the district's 2008 accountability rating was academically acceptable (TEA, 2008e).

Participants

The sample for my study consisted of 185 Hispanic ELLs (n = 185) from the Project ELLA study, which is considered an adequate sample size for SEM analysis (Thompson, 2000). Slightly more than one half of the students were male (53.5%) while 46.5% were female. Students received instruction in one of four instructional models in the Project ELLA study: SEI-E. SEI-T, TBE-E, and TBE-T. The number of participants in each of the four groups ranged from 74 to 18 students. Table 5 shows the distribution of students with no missing data across study conditions and the language version of the TAKS reading test. The entire sample of third grade Hispanic ELL Project ELLA participants with no missing data who took the English and Spanish TAKS test was 321 students. Only Project ELLA participants who took the English TAKS test were examined in this study.

Table 5.

Distribution of Participants Among the Four Study Conditions and TAKS Test Language Version Administered

	% of Sample	English TAKS [a]	% English TAKS[b]	Spanish TAKS [c]	% Spanish TAKS[d]	Project ELLA Cohort
SEI-E (*n* = 54)	29.2	54	100.0	0	00.0	100.0%
SEI-T (*n* = 74)	40.0	74	100.0	0	00.0	100.0%
TBE-E (*n* = 39)	21.1	39	36.1	69	63.9	100.0%
TBE-T (*n* = 18)	9.7	18	21.2	67	78.8	100.0%
Total (*n* = 185)	100.0	*n*=185		*n*=136[e]		*n*=321

Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control. [a]Participants in each program model who took the English TAKS reading test. [b]Percentage of total number of students in each study condition who took the English TAKS reading test. [c]Students from transitional bilingual education experimental and control conditions who took the Spanish TAKS reading test. [d]Percentage of total students from transitional bilingual education experimental and control conditions who took the Spanish TAKS reading test. [d]Total number of students from the bilingual education experimental and control conditions who had no missing data but were not included in my study because only English oral language trajectory as a predictor of English TAKS reading was examined.

Measure

Oral language and reading achievement performance were operationalized using the PV and the LC subtests of the WLPB-R English Form (Woodcock, 1991) and the third grade ETAKS reading test. The PV and the LC subtests measured oral language trajectory over 5 time points from kindergarten through third grade. While the PV and LC subtests share a single word answer format, they differ in the complexity of the language skills required of the examinee. A second distinction is the language skill task modality, whether the modality is primarily *receptive* or *expressive*.

WLPB-R picture vocabulary and listening comprehension subtests. The PV subtest, which consists of 58 items, is a word retrieval and expressive semantic task (Woodcock, 1991b). WLPB-R subtests represent a range of language skills from simple to complex. The PV subtest falls at the lower end (simple) of this complexity range. In contrast the LC subtest, consisting of 38 items, is primarily a receptive language test. As *connected discourse*, this task falls within the mid level of the WLPB-R language task complexity range. Connected discourse means that the item stimulus or cue is not presented in isolated units but rather "calls upon previously acquired knowledge and the ability to conclude or predict based on the information presented" (Woodcock, 1991b; p. 74). The item stimulus is usually a passage with an oral *cloze* response (Woodcock, 1991b). Reliability evidence for the WLPB-R PV and LC subtests are presented in the next section.

PV and LC subtests item scores from the secondary data were available for the following years with a Cronbach's alpha, an internal consistency estimate, as follows:

.78 for PV 2006-2007; .66 for PV 2007-2008; .80 for LC 2006-2007; .72 for LC 2007-2008. This reliability evidence is supported by reliability indexes reported in the WLPB-R manual (Woodcock, 1991b). Statistics were reported in the manual at nine age group levels, ranging from 2 – 79 years. Based on a split-half procedure, reliability coefficients (internal consistency) reported for PV subtest data for individuals, which correspond to the participant ages of the longitudinal data in my study, were .773 at age 6 and .845 at age 9. For LC subtests data, reliability coefficients reported were .826 at age 6 and .810 at age 9 (Woodcock, 1991b). Construct validity statistics were reported as a correlation statistic between individual subtests and cluster scores. Statistics reported for the PV subtest and the Reading Comprehension Cluster data were .278 (age 6) and .610 (age 9). LC subtest and Reading Comprehension Cluster data yielded correlation statistics of .338 (age 6) and .612 (age 9; Woodcock, 1991b).

English TAKS reading test. The third grade ETAKS reading test is a standards-referenced assessment for the Texas Essential Knowledge and Skills (TEKS). This reading comprehension test consists of 36 multiple-choice items. According to the Texas Education Agency (TEA) *Technical Digest 2006-2007,* internal consistency was calculated using the Kuder-Richardson Formula 20 (KR20) and ranged from .83 to .93. Validity coefficients for the TAKS test data were not reported. Instead, evidence was provided that content validity for TAKS test data is "content based and tied directly to the statewide curriculum" (TEA, 2006, p. 177).

Data Analysis

Structural equation modeling techniques were employed for fitting a latent growth model

to the longitudinal data. “One of the unique and powerful advantages of utilizing LGM methodology over traditional methods is its ability to incorporate predictors of the latent growth factors, thereby attempting to explain individual differences in latent trajectories” (Hancock & Lawrence, 2006; p. 187). Successfully fitting a latent growth model (LGM) to longitudinal data, using structural equation modeling techniques, captures the change process: how it occurs, how much occurs, and how it differs across the individual study participants. (Hancock & Lawrence, 2006).

Two kinds of statistical software were used in my study. SPSS (Version 16.0) and Multinor (Henson, 1999; Thompson, 1990) were used to perform descriptive, univariate, and multivariate analyses; to assess group differences; and to evaluate univariate and multivariate normality. A multivariate statistical analysis was performed using a latent variable modeling technique to test the hypothetical model fit, to examine the relationship between the predictor and outcome variables, and to assess the impact of instructional model group membership on individual participant oral language outcomes. Based on the characteristics of the data, a structural equation model (SEM) approach was used to test the model. Once the computer program estimated the model, the following protocol was followed as suggested by Kline (2005):

- The model fit was evaluated;
- The parameter estimates were interpreted; and
- Equivalent models were considered.

Mplus (Version 5.1) was the statistical software program used to analyze the sample data set. To analyze the longitudinal data, my study employed a latent growth

model with structural equation modeling techniques as recommended by Muthén and Muthén (2007). A latent growth model for this longitudinal data set with 5 time points was estimated. The structural equation modeling approach produced time score parameters, which yielded the growth function estimates. The effect of group membership was measured using dummy codes (Muthén & Muthén, 2007). SEI-T was the reference group for dummy coding. Maximum likelihood (ML) estimator was used to estimate the model's parameters.

Results

Descriptive Statistics

In preparation for the statistical model analysis, data were screened as suggested by Kline (2005) to identify issues relating to multivariate normality, one of the assumptions of ML estimator. Univariate normality is a necessary condition of multivariate normality but not a guarantee (Stevens, 2002; Thompson, 2006). Univariate statistics normality tests, Shapiro-Wilk and Kolmogorov Smirnov, were statistically significant for all observed variables with the exception of Listening Comprehension, time 4. However, the coefficients of skewness for all observed variables were no higher than 1 and the coefficients of kurtosis were no higher than 2. Finney and DiStefano (2006) pointed out that while there is no acceptable degree of non-normality, some studies suggested that only when skewness coefficients approach 2 and kurtosis coefficient approach 7 are there problems reported with ML estimator results. Loehlin (2004) and Stevens (2002) suggest the importance of examining the scatterplots for each pair-wise combination. Scatterplots and a Multinor graph were examined to identify outliers that might contribute to non-normality. Based on visual inspection, all data

distributions were generally elliptical and followed the Multinor normal distribution reference line fairly well.

Table 6 presents the mean and standard deviations of the *WLPB-R* PV and LC subtests and the ETAKS. The line graph in Figure 6 presents the 5 time point means of the PV subtests. The results showed a statistically significant difference in the beginning English proficiency levels among the four study conditions or program models. A similar trajectory pattern is evident in the line graph of the LC subtests shown in Figure 7. The bar graph in Figure 8 presents the mean TAKS scores of the study conditions. The difference in the scores is quite interesting since group membership was a significant factor in only the TBE-T (control) group. In the hypothetical model this corresponds to the path from group to TAKS. While this was not one of the research questions posed in my study, the difference in ETAKS scores does merit comment. One possible explanation is that this statistically significant difference is a teacher effect. The TBE-T group was the smallest of the four program models with 18 students. It appears that almost all of the participants in TBE-T were taught by two teachers. Sixteen of the 18 students were evenly divided between two teachers albeit at different schools.

Table 6.

Means and Standard Deviations of Predictor Variables, Picture Vocabulary[a] and Listening Comprehension[b] Subtest Scores and Outcome Variable, TAKS[c] Scores

		PV 1	LC1	PV 2	LC2	PV 3	LC3	PV4	LC4	PV 5	LC5	TAKS
SEI-E	*M*	19.35	5.98	22.65	9.13	24.13	14.09	27.24	18.07	29.11	20.50	2284.09
	SD	3.16	4.37	3.04	4.79	2.88	4.31	3.30	4.05	2.35	3.25	150.14
SEI-T	*M*	21.23	7.59	23.09	10.41	25.57	15.39	28.31	18.35	29.81	20.62	2292.69
	SD	3.53	4.48	2.72	4.46	2.85	3.81	3.20	4.05	2.69	3.21	154.36
TBE-E	*M*	14.46	2.82	18.67	4.31	21.95	10.67	25.31	15.33	28.10	18.95	2262.31
	SD	4.82	3.28	2.56	3.54	3.24	4.60	3.35	4.00	2.75	3.98	117.69
TBE-T	*M*	14.5	2.67	19.11	6.00	21.94	12.11	24.56	14.67	28.44	19.56	2363.44
	SD	3.70	2.35	2.54	4.19	3.40	5.04	4.25	3.80	3.07	2.77	160.25
Total	*M*	18.60	5.64	21.64	8.32	24.03	13.70	27.00	17.28	29.11	20.13	2290.66
	SD	4.70	4.51	3.35	4.96	3.34	4.61	3.62	4.24	2.71	3.40	147.93

Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control; PV = Picture Vocabulary subtest of the Woodcock Language Proficiency Battery-Revised; LC = Listening Comprehension subtest of the Woodcock Language Proficiency Battery-Revised; 1 = data collected in 2004; 2 = data collected in 2005; 3 = data collected in 2006; 4 = data collected in 2007; 5 = data collected in 2008. [a]Picture Vocabulary subtest consists of 58 items with raw score range between 0-58. [b]Listening Comprehension subtest consists of 38 items with raw score range between 0-38. [c]TAKS reading tests consists of 36 items with scale score range between 1875-2615.

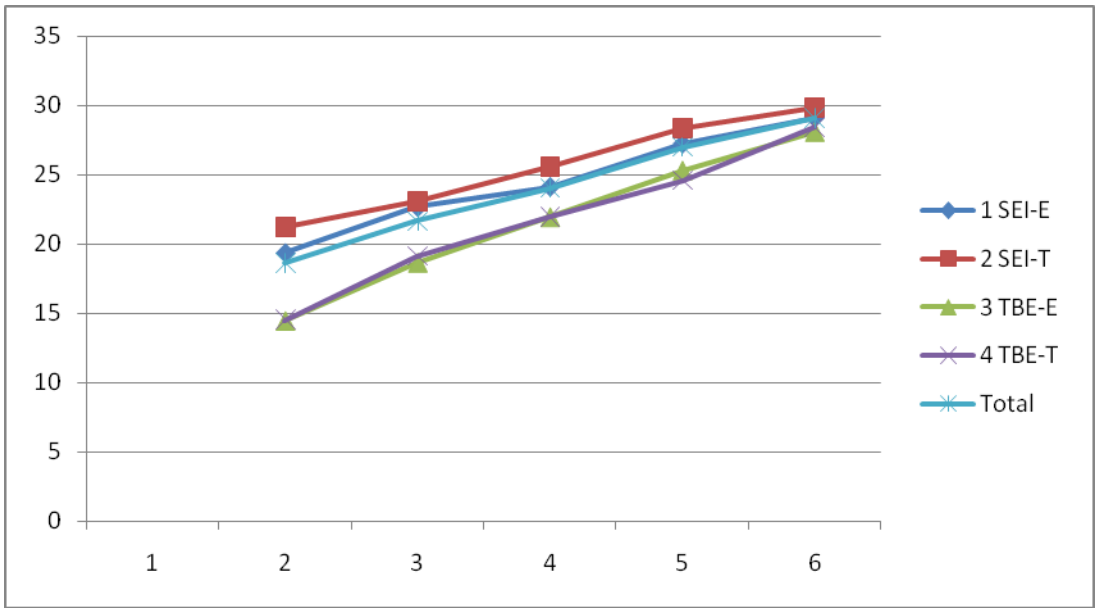

Figure 6. Mean Scores of Picture Vocabulary Subtests for Five Time Points.

Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control.

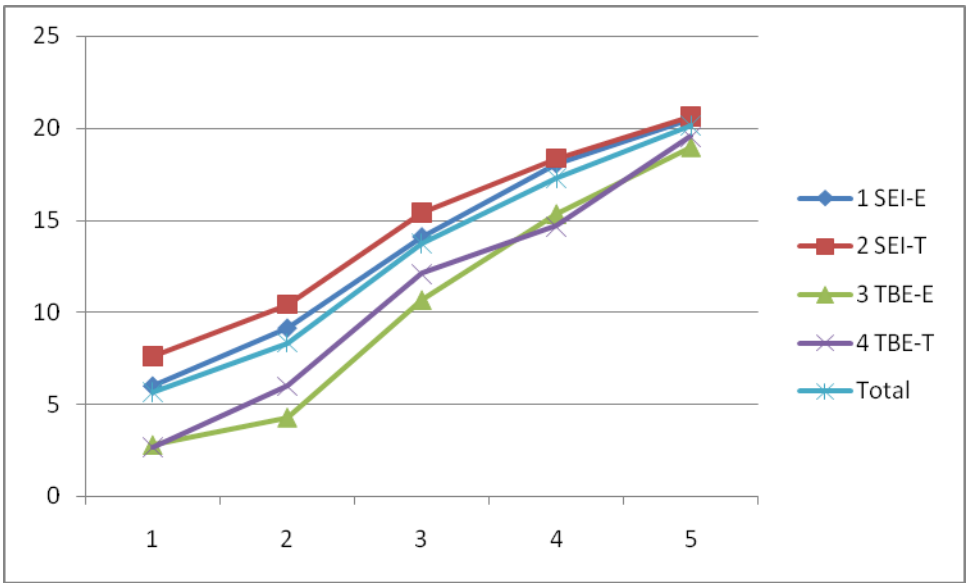

Figure 7. Mean Scores of Listening Comprehension Subtests for Five Time Points.
Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control.

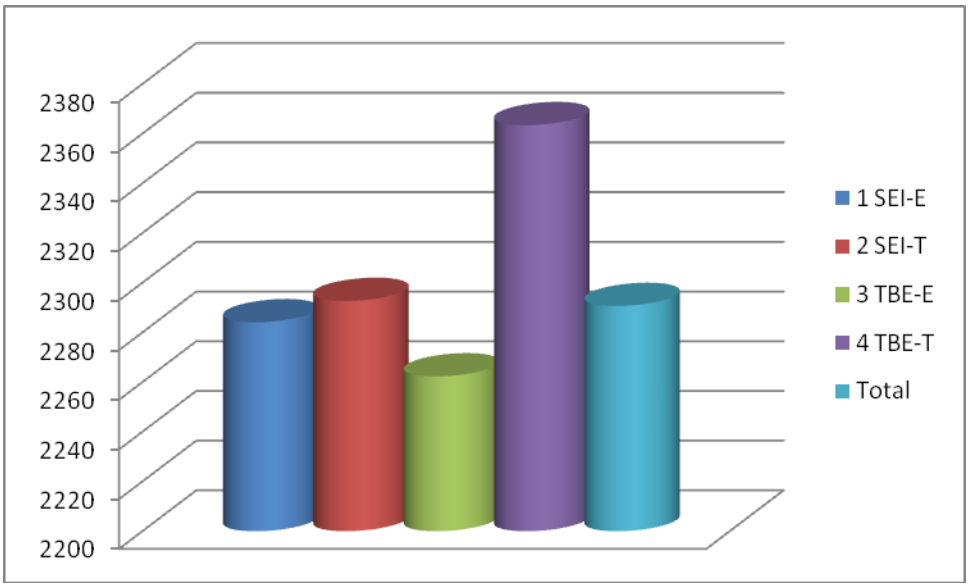

Figure 8. English TAKS Reading Mean Scores of Four Program Models.
Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control.

Model Results

The final model for PV produced an R^2 statistic of 24.6%, 95% CI .142-.350 with an adjusted R^2 statistic of 22.5%, 95% CI .123-.327, displayed in Table 7. Table 8 shows the final model for LC, which produced an R^2 statistic of 32.1%, 95% CI .214-.428, with an adjusted R^2 30.2%, 95% CI .195-.409. The R^2 statistic summarizes the goodness of fit of the model to the longitudinal data (Singer & Willett, 2003). In addition, R^2 indicates the strength of the oral language variables PV and LC as predictors of the outcome variable third grade ETAKS reading scores. Therefore this latent growth model for PV and LC explained from 22% to 30% of the variability of the third grade ETAKS reading performance for the Hispanic ELL participants.

A second goodness-of fit statistic, chi-square, is listed in Tables 7 and 8 along with the following fit indexes: Comparative Fit Index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Based on a null hypothesis testing approach, the chi-square statistic, as pointed out by Hu and Bentler (1998), "may not be a reliable guide to model adequacy" (p. 425). For this reason, fit indexes, which quantify a model fit to the data along a continuum, are also commonly reported. For the PV latent growth model the fit indexes were CFI (.970), RMSEA (.075), and SRMR (.095). The LC latent growth model fit indexes were CFI (.915), RMSEA (.075), and SRMR (.063). According to the cutoff criteria for fit indexes recommended by Hu and Bentler (1999), the PV latent growth model is a good to poor

description of the data.[10] Using these same cutoff criteria, the LC latent growth model does not describe the data quite as well. In response to these *rule of thumb* fit index cutoff criteria, other studies (Fan & Sivo, 2005, 2007; Fan, Thompson, & Wang, 1999; Marsh, Hau, & Wen, 2004) report that model fit indexes can vary across different conditions and sample sizes and therefore are not absolute values of adequate theoretical model fit to the empirical data. Furthermore, Klein (2005) defines a CFI value above .90 as a reasonable good fit. Finally, Hu and Bentler (1998) suggest that model validity is established by the examination of fit indexes, the overall interpretability of the correlational coefficients (i.e., the parameter estimates), model complexity, and previous studies in the research area. The remaining results are organized and reported by the two research questions of my dissertation study.

Research Question Number 1

The first research question was: Does the English oral language proficiency trajectory of Hispanic ELLs explain their performance on a third grade state-wide high-stakes reading test in English?

[10] Hu and Bentler (1999) suggest when using ML as the estimator fit index cutoff values should be near to the following:.95 for CFI, ..08 for SRMR, and .06 for RMSEA.

Table 7.

Latent Growth Model Summary for Picture Vocabulary Subtest Scores[a]

Parameter	Unstandardized	SE	Standardized	SE	*p*-value
Mean - I	21.033	.349	5.588	.336	<.001
Mean - S	2.215	.088	2.833	.318	<.001
I ↔ S	-1.022	.292	-.589	.074	<.001
I → TAKS	.295	.046	.753	.110	<.001
S → TAKS	1.023	.249	.542	.122	<.001
D1 → I	-1.120	.544	-.135	.066	.040
D3 → I	-5.943	.596	-.644	.058	<.001
D4 → I	-5.688	.792	-.448	.061	<.001
D1 → S	.097	.136	.057	.079	.476
D3 → S	1.064	.149	.555	.079	<.001
D4 → S	1.052	.199	.399	.076	<.001
D1→TAKS	.145	.237	.045	.073	.541
D3→TAKS	.362	.355	.100	.098	.308
D4→TAKS	1.311	.415	.263	.082	.001
R^2 - TAKS	.246	.060			<.001
R^2 - S	.365	.082			<.001

Note. I = Intercept; S = Slope; SE = Standard Error; TAKS = English TAKS reading test. D1, D3, D4 are dummy coding groups. The reference groups is SEI-T structured English immersion – typical/control. D1 = SEI-E structured English immersion – enhanced/experimental; D3 = TBE-E transitional bilingual education - enhanced/experimental; D4 = TBE-T transitional bilingual education - typical/control.
[a]Fit Indices: $\chi 2$ 41.038 (p = .004); CFI, .970; RMSEA, .075; and SRMR, .095.

Table 8.

Latent Growth Model Summary for Listening Comprehension Subtest Scores[a]

Parameter	Unstandardized	SE	Standardized	SE	*p*-value
Mean - I	7.703	.446	1.978	.145	<.001
Mean - S	3.271	.114	4.680	.756	<.001
I ↔ S	-.859	.390	-.468	.119	<.001
I → TAKS	.275	.039	.726	.104	<.001
S → TAKS	1.064	.343	.504	.144	.002
D1 → I	-1.700	.684	-.199	.079	.013
D3 → I	-5.668	.768	-.594	.072	<.001
D4 → I	-5.121	1.004	-.390	.073	<.001
D1 → S	.400	.172	.260	.113	.020
D3 → S	.903	.194	.527	.120	<.001
D4 → S	.924	.253	.392	.108	<.001
D1→TAKS	-.044	.261	-.014	.080	.866
D3→TAKS	.295	.383	.082	.106	.441
D4→TAKS	1.134	.440	.228	.088	.010
R^2 - TAKS	.321	.071			<.001
R^2 - S	.295	.116			.011

Note. I = Intercept; S = Slope; SE = Standard Error; TAKS = English TAKS reading test. D1, D3, D4 are dummy coding groups. The reference group is SEI-T structured English immersion – typical/control. D1 = SEI-E structured English immersion – enhanced/experimental; D3 = TBE-E transitional bilingual education - enhanced/experimental; D4 = TBE-T transitional bilingual education - typical/control.
[a]Fit Indices: $\chi 2$ 64.964 (p = <.001); CFI, .915; RMSEA, .110; and SRMR, .063.

Picture vocabulary model. Table 7 summarizes the latent growth model parameters for the PV subtest model. The intercept (B=.753; $p<.001$) and the slope (B=.542; $p<.001$) were statistically significant predictors of future ETAKS reading scores. The magnitude of this relationship was expressed in a comparison of the factor pattern coefficients, which reflected an intercept factor pattern coefficient that was higher than the slope factor pattern coefficient. Thus, while initial English proficiency (intercept) and English language acquisition rates (slope) over the 5 time period predicted ETAKS reading performance, the English that Hispanic ELLs knew when they entered kindergarten was a somewhat better predictor than the English students learned as they progressed from grades K-3, i.e. the oral language trajectory. In sum, Hispanic ELLS who were more proficient in English at the start of kindergarten had a high probability of scoring more points on the ETAKS reading test.

The next parameter to consider is the relationship between the intercept and the slope. Table 9 shows the covariance matrix for the PV latent growth model. The covariance statistic in Table 7 is negative (-.589, $p<.001$) and statistically significant. A negative covariance indicates that when values fall above the mean of the intercept they tend to have a slope or oral language trajectory that is below the mean of the slope. Likewise, when values fall below the intercept mean they tend to have a slope or oral language trajectory that is above the mean of the slope. Consequently Hispanics ELLs with higher initial proficiency tended to have a lower English acquisition rate that those Hispanic ELLs with lower initial English proficiency levels who tended to have higher

Table 9.

Covariance Matrix of Picture Vocabulary Subtest Growth Model

	PV 1	PV 2	PV 3	PV 4	PV 5	TAKS	D1	D3	D4
PV 1	22.002								
PV 2	10.928	11.127							
PV 3	10.532	7.963	11.080						
PV 4	10.389	7.497	9.124	12.995					
PV 5	6.240	4.808	5.840	6.659	7.311				
TAKS	1.634	1.063	1.468	1.684	1.634	2.176			
D1	0.219	0.293	0.028	0.070	-0.001	-0.019	0.207		
D3	-0.872	-0.627	-0.439	-0.357	-0.213	-0.060	-0.062	0.166	
D4	-0.399	-0.246	-0.203	-0.238	-0.065	0.071	-0.028	-0.021	0.088

Note. PV = Picture Vocabulary subtest of the Woodcock Language Proficiency Battery-Revised; TAKS = English TAKS reading test; 1 = data collected in 2004; 2 = data collected in 2005; 3 = data collected in 2006; 4 = data collected in 2007; 5 = data collected in 2008. D1, D3, D4 are dummy coding groups. The reference group is SEI-T structured English immersion – typical/control. D1 = SEI-E structured English immersion – enhanced/experimental; D3 = TBE-E transitional bilingual education - enhanced/experimental; D4 = TBE-T transitional bilingual education - typical/control.

English acquisition rates. Students showed 2.215 (p<.001) points growth per year on the PV subtest. The LC latent model will be considered next.

Listening comprehension model. The latent growth model parameters for the LC subtest model are summarized in Table 8. The intercept (B=.726; *p*<.001) and the slope (B=.504; *p* = .002) were statistically significant predictors of future ETAKS reading scores. Like the PV factor pattern coefficients, the intercept factor pattern coefficient was higher than the slope factor pattern coefficient and was also a somewhat better predictor of prospective ETAKS reading performance. The next parameter to be examined is the correlation between the LC intercept and slope. Table 10 shows the covariance matrix for the LC latent growth model. The correlation between LC intercept and slope shared the same direction as that of the PV covariance parameter. The negative covariance (-.468, *p*<.001), shown in Table 8 signifies an inverse relationship between beginning English proficiency and English acquisition rates. Students showed 3.271 (*p*<.001) points growth per year on the LC subtest. Finally, the relationship between observed variables, displayed in Table 11, shows a rank order that appears to remain stable over the 5 time periods. While the correlations are all statistically significant (*p*<.001), they ranged from .236 (ETAK to PV1) to .733 (PV4 to LC4). The second research question considered the effect of group membership.

Research Question Number 2

The second research question was: Does ELLs' instructional model assignment explain their English oral language trajectory?

Picture vocabulary model. The factor pattern coefficients of the intercept for all three study conditions, presented in Table 7, were negative and statistically significant when compared to the reference group, SEI-T. The negative factor pattern coefficients indicate the English proficiency levels of the students in the SEI-T were significantly higher than the English proficiency levels of the other three study conditions when the students entered kindergarten. The students in the TBE experimental and control conditions had the lowest beginning English proficiency levels at $p<.001$. In contrast, the factor pattern coefficients for the oral language trajectory were statistically significant for TBE-E (.555, $p<.001$), and TBE-T (.399, $p<.001$) but not for SEI-E. Moreover, the oral language trajectory for each of the three study conditions was positive in relation to the reference group. Thus, while the Hispanic ELLs in SEI-T as a group knew more English when they began kindergarten, the oral language trajectory or acquisition rate for each of the other three groups was steeper, even more so for the TBE experimental and control conditions.

Table 10.

Covariance Matrix of Listening Comprehension Subtest Growth Model

	LC 1	LC 2	LC 3	LC 4	LC 5	TAKS	D1	D3	D4
LC 1	20.188								
LC 2	14.024	24.466							
LC 3	11.820	13.356	21.108						
LC 4	9.721	12.555	12.127	17.897					
LC 5	7.058	7.251	8.823	9.007	11.507				
TAKS	1.499	2.386	2.205	2.504	2.134	2.176			
D1	0.100	0.237	0.115	0.233	0.108	-0.019	0.207		
D3	-0.594	-0.846	-0.639	-0.409	-0.249	-0.060	-0.062	0.166	
D4	-0.289	-0.226	-0.154	-0.254	-0.056	0.071	-0.028	-0.021	0.088

Note. LC = Listening Comprehension subtest of the Woodcock Language Proficiency Battery-Revised; TAKS = English TAKS reading test; 1 = data collected in 2004; 2 = data collected in 2005; 3 = data collected in 2006; 4 = data collected in 2007; 5 = data collected in 2008. D1, D3, D4 are dummy coding groups. The reference groups is SEI-T structured English immersion – typical/control. D1 = SEI-E structured English immersion – enhanced/experimental; D3 = TBE-E transitional bilingual education - enhanced/experimental; D4 = TBE-T transitional bilingual education - typical/control.

Table 11.

Correlations Among Observed Variables

	PV1	LC1	PV2	LC2	PV3	LC3	PV4	LC4	PV5	LC5	TAKS
PV1	–										
LC1	.670**	–									
PV2	.698**	.555**	–								
LC2	.606**	.631**	.585**	–							
PV3	.675**	.563**	.717**	.585**	–						
LC3	.603**	.573**	.588**	.588**	.663**	–					
PV4	.614**	.571**	.624**	.567**	.760**	.583**	–				
LC4	.518**	.511**	.545**	.600**	.624**	.624**	.733**	–			
PV5	.492**	.503**	.533**	.478**	.649**	.555**	.683**	.593**	–		
LC5	.436**	.463**	.389**	.432**	.505**	.566**	.539**	.628**	.536**	–	
TAKS	.236*	.226*	.216*	.327**	.299**	.325**	.317**	.401**	.410**	.426**	–

Note. SEI = structured English immersion; TBE = transitional bilingual education; E = enhanced/experimental; T = typical/control; PV = Picture Vocabulary subtest of the Woodcock Language Proficiency Battery-Revised; LC Listening Comprehension subtest of the Woodcock Language Proficiency Battery-Revised; TAKS = English TAKS reading test; 1 = data collected in 2004; 2 = data collected in 2005; 3 = data collected in 2006; 4 = data collected in 2007; 5 = data collected in 2008. *$p < .05$. **$p < .001$.

Listening comprehension model. An examination of the LC factor pattern coefficients computed using the dummy codes to compare study conditions revealed results similar to those of the PV latent growth model. All three study conditions had negative factor pattern coefficients that were statistically significant when compared to the factor pattern coefficient of SEI-T. The oral language trajectory for each of the study conditions was positive and statistically significant when compared to the reference group. Therefore, while the students as a group in the SEI-E, TBE-E, and TBE-T experimental and control conditions had lower beginning English proficiency levels than the SEI-T group, their learning rates were steeper, meaning that the rate of English language acquisition for each of these study conditions was higher than English acquisition rate of the SEI-T control condition.

Discussion

My study investigated the relationship between L2 oral language and L2 reading comprehension outcomes for Hispanic ELLs as measured by the third grade ETAKS test. The two discrete indicators of English oral language proficiency were vocabulary knowledge and listening comprehension. Therefore, the PV and LC subtests of the WLPB-R were employed as proxy measurements of oral language.

The two theoretical models (PV and LC) fit the longitudinal data reasonably well and accounted for a unique variability of 23% and 30%, respectively. These findings, which identify vocabulary knowledge and listening comprehension as two component skills for facile reading comprehension, support similar evidence from previous research studies of ELL and English-only (EO) students: Hispanic ELLs (August, Carlo, Dressler,

& Snow, 2005; August et al., 2006b; Proctor et al., 2005; Tong et al., 2008), ELLs from multiple linguistic backgrounds (Geva & Zadeh, 2006; Pajoohesh, 2007; Verhoeven, 2000), and EO students (Carver & David, 2001; Storch & Whitehurst, 2002).

The change in EL2 acquisition over the five time points was marked by a strong positive trajectory in both latent growth models. Models contained annual growth rates of slightly over two points for PV and three points for LC. Differences among the initial English oral language proficiency levels and growth rates of individual ELLs were statistically significant. Students with the lowest initial EL2 oral language proficiency levels tended to have the steepest learning rates, while those with the highest initial EL2 oral language proficiency levels tended to have the flattest learning rates. The best predictors of ETAKS reading outcomes were English proficiency levels at the beginning of kindergarten followed by English acquired over the 4-year period from kindergarten through the third grade.

The second relationship investigated in my study concerned instructional programs and their effect on oral language acquisition. Groups were dummy coded with SEI-T as the comparison group. Language of instruction (TBE-E and TBE-T), i.e., program models with English and Spanish instruction, was statistically significant in the PV latent model. In the LC latent growth model, language of instruction and the experimental treatment program model (SEI-E) were statistically significant. Thus, instruction in ELLs' primary language, i.e., Spanish, appears to support English acquisition for Spanish-speaking ELLs. These findings are congruent with the research findings of Tong et al. (2008) who examined the growth trajectory and the EL2 oral

language acquisition of Project ELLA study participants from kindergarten through the end of the first grade. As a group, Hispanic ELLs in SEI-T knew the most English when they began kindergarten followed by SEI-E and the TBE groups. The TBE groups' growth rates were steeper than SEI-T in verbal knowledge and listening comprehension while the SEI-E growth rate was clearly steeper than SEI-T in listening comprehension. Thus, students in the program models with steeper growth rates experienced greater increases in their verbal knowledge and listening comprehension.

An interesting result was the relationship noted between instructional programs and ETAKS performance. TBE-T was the only study condition where instructional program predicted ETAKS outcomes. While this association may be a teacher effect, other possible explanations are class size and instructional time allotment for Spanish and English. A characteristic of bilingual program models can be the instructional time allotment of primary and secondary languages. Finally, Irby et al. (2007) notes that the actual language distribution in content areas of individual classrooms may not reflect district guidelines.

High-stakes testing impacts Hispanic ELLs individually and as a group, vis-à-vis grade retention and school completion rates (Solórzano, 2008). The strengths of my study are threefold. First, in this study the association between oral language and reading comprehension performance using longitudinal data from a quasi-experimental study spanning a 4-year period from kindergarten-third grade was examined. Thus, evidence from this study supports similar findings establishing the causal link between EL2 oral language and EL2 reading comprehension in both ELLs and EO students that appear in

published research literature (August, 2003; Carver & David, 2001; Proctor et al., 2005; Storch & Whitehurst, 2002). Second, the proxy for reading comprehension is a large-scale high-stakes reading test, i.e., the ETAKS reading test. Few research studies have examined high-stakes testing and ELLs (Mahon, 2006). Third, this study examined instructional programs. Findings from this study appear to support the importance of language of instruction (TBE-E and TBE-T), i.e. ELLs' primary language, and EL2 instructional methodologies (SEI-E and TBE-E) employing evidence-based best practices in language and literacy acquisition for ELLs.

Finally, the pressing issues in ELL education today are linguistic competence in English and academic parity with non-ELL peers. To address these issues more research evidence is needed regarding the relationship between oral language and other domain literacies, e.g. math and science. Furthermore, studies have generally focused on the primary grades. A critical demographic for future research are secondary ELLs. Funding longitudinal studies for the upper elementary grades through the secondary grades could contribute to our knowledge of the discrete stages of EL2 language development. In addition, evidence based strategies and instructional models may be identified that enhance and accelerate EL2 acquisition rates from very young to adolescent ELLs. As a final consideration, analytical techniques that maximize our ability to quantify and contextualize our findings should be considered in future research agendas. Mixed-methods research is a promising paradigm that accomplishes both of these goals.

Assumptions

My study included the following assumptions related to test administration and instructional delivery in study two, the quantitative study. These assumptions are based on Project ELLA program descriptions submitted to the Texas A&M University Office of Research Compliance and personal communications with Project ELLA coordinators and principal investigators (PIs; May 2007).

1. *WLPB-R* PV and LC subtests and the ETAKS reading test were administered by trained personnel and all testing procedures specific to each test as defined by the *WLPB-R* manual and the Student Assessment Division of the Texas Education Agency were followed.
2. All classroom teachers who participated in Project ELLA were certified bilingual and/or ESL teachers who have fulfilled the Texas State Board for Educator Certification (SBEC) certification requirements.

Limitations

A limitation of this study should be acknowledged, which may limit the generalizability of the study findings. Oral language is an array of skills that includes but is not limited to verbal knowledge and listening comprehension. In this study oral language was operationalized as verbal knowledge and listening comprehension. Therefore, other domains of oral language, such as grammar or narrative discourse, were not included.

Implications for Practice

There are measurement and pedagogical implications of my research study. Findings from this study confirm that EL2 oral language may explain unique variability in the ETAKS reading test. These findings are consistent with evidence from research syntheses of English literacy and ELLs (August, 2003; Genesee, Lindholm-Leary, Saunders, & Christian, 2006). The causal link establishing verbal knowledge and listening comprehension as predictors of reading comprehension levels suggests the existence of developmental properties in the EL2 oral language acquisition process. Defining these properties would allow researchers to refine EL2 oral language development theories and to specify EL2 oral language acquisition stages with their respective ranges. Eventually researchers might be able to establish parameters of normal, delayed, and impaired EL2 oral language functioning (Jean & Geva, 2009). To accomplish this goal, educators need more precise assessment instruments for measuring EL2 reading comprehension and student EL2 oral language production. Such reading comprehension assessments would assess those underlying skills of reading comprehension, including oral language, thus providing practitioners with an individualized diagnostic profile. This would allow practitioners to more effectively differentiate reading instruction (August, Francis, Hsu, & Snow, 2006a). Data from measures of ELLs' oral language production would establish EL2 oral language acquisition benchmarks and inform pedagogy.

Citing research findings from their longitudinal study of Hispanic ELL literacy development in elementary school, August and her colleagues (2006) discussed second

language learning pedagogy and pointed out the importance of robust vocabulary instruction, quality language learning environments, and awareness of second language learning as an uneven course given the number of factors involved. My study has implications for pedagogy that fall across four quadrants: oral language development strategies, teacher practices and preparation, pre-kindergarten and pre-school programs, and home literacy.

Considering the first quadrant, what are effective, evidence-based EL2 oral language development strategies? The keystone strategy is vocabulary development. Ordóñez, Carlo, Snow, & McLaughlin (2002) studied depth and breadth of vocabulary development in L1 and L2 and identified vocabulary depth knowledge as crucial to the development of decontextualized oral language skills. In order to increase word depth knowledge, ELLs must be cognizant of semantic networks related to each word, its syntactic structures, its phonological and orthographic representations, and its morphological structure. Other studies have evidenced this association between increased domain word knowledge and instructional models that incorporate strategies for identifying Spanish-English cognates and English-Spanish as the languages of instruction for Hispanic ELLs (García, 1991; Nagy, García, Durgunoğlu, & Hacin-Bhatt, 1993). Similarly additional studies (Carlo et al., 2004; Duursma et al., 2007; Proctor et al., 2005) have indicated the importance of explicit instruction rather than incidental learning approaches as crucial to English language acquisition for ELLs.

A second effective evidence-based instructional approach that increases students' vocabulary knowledge is instructional conversation, a classroom discourse technique

described by Padrón and Waxman (1999). Their findings demonstrate a correlation between instructional conversation and increased student reading comprehension. Knight & Wiseman (2006) identified instructional conversations as an effective instructional approach for ELLs. Instructional conversations appear to contextualize reading instruction through the lens of the students' lives, as individuals, in their families and communities. Thus, the instructional conversation approach "explicitly targets formulation and expression of ideas through oral language and can be used prior to, during, and after reading and in combination with other approaches" (Knight & Wiseman, 2006; p. 81).

The second quadrant addresses teacher practices and preparation. How do these EL2 oral language strategies concord with research findings on effective teacher practices and preparation? Calderón (2006) identified the most important professional competency to improve teacher performance as knowing how to teach reading to ELLs, e.g. research-based literacy instruction. Literacy in both languages and "when and how to teach L1 and L2" (Calderón, 2006; p. 125) are components that should be integrated in all ELL teacher repertoires, in teacher preparation programs, and in professional development plans. Slavin and Cheung (2005) conducted a best evidence research synthesis in which 13 of the 17 qualifying studies examined reading instruction and language of instruction among Spanish-speaking ELLs. Based on their findings, the authors reported that 70% of the 13 studies favored instruction in L1 and L2. In the present study, language of instruction (i.e., English/Spanish) and the structured English

immersion intervention emerged as important instructional models in EL2 oral language development.

While literacy development for ELLs is a critical component of teacher repertoires, Wong Fillmore and Snow (2002) stressed that teacher education should also include mastery of specific knowledge about language and the linguistic components of oral language. Their findings underscore the need for bilingual education teacher preparation programs at the baccalaureate level to offer rigorous coursework in second language acquisition theory, biliteracy, domain knowledge in the content areas, and rigorous fieldwork in bilingual classrooms. Pre-service teachers should be partnered in the field with competent and articulate bilingual teachers who are willing to engage in dialectical conversations about their practice. For masters programs, learning outcomes should center on teacher instructional leadership, mentoring, and critical pedagogy that incorporates a research literacy strand. Last of all, professional development for in-service teachers should provide strategies and opportunities to improve peer collaboration and coaching skills, to design inquiry lessons that build conceptual knowledge and incorporate differentiated instruction, and to practice instructional delivery strategies that enhance academic vocabulary knowledge and instructional discourse.

The best predictor of ETAKS reading outcomes was the initial EL2 verbal knowledge and listening comprehension proficiency levels of Hispanic ELLs when they entered kindergarten. Thus, the third quadrant is encapsulated in the following question: What are the characteristics of effective preschool programs? Snow (1991) listed four

domains of school readiness language skills that should be addressed in pre-school curricula and supported by instructional methodology and strategies. These domains include conversational language skills, decontextualized oral language skills, print skills, and emergent literacy skills. Additionally, Rolla San Francisco, Carlo, August, and Snow (2006) found that Spanish phonological skills did transfer to English vocabulary in unbalanced[11] bilingual kindergarten and first grade students receiving Spanish and English instruction. This transference has in general been supported by the research literature and bilingual education theory (García & Jensen, 2007). Thus, biliteracy development for ELLs should begin with all day preschool programs for 3 year olds. The programs should be located in primary schools where preschool teachers can collaborate within and across grade levels. Instruction should be guided by evidence-based curricular frameworks that emphasize verbal knowledge and listening comprehension in L1 and L2. Finally a parent involvement and parent training strands should be in place to support and increase home literacy practices.

Last of all, the fourth quadrant asks, how do home literacy practices impact ELLs' literacy development? Research suggests that home literacy practices correlate to Hispanic children´s language development and should be targeted in early childhood programs (Jordan, Snow, & Porche, 2000). Thus, early childhood programs should offer parent workshops in which instructors model reading to young children. These workshops would provide opportunities for parents to read to their children. Home

[11] By definition ELLs are unbalanced bilinguals, meaning that the child's is proficient in L1 but not in L2, the target language.

libraries, in addition to audio books, are other resources that preschool programs can incorporate into their reading curricula. A second aspect of home literacy training is teaching parents questioning strategies that support oral language development and promote conversations or dialogues about books (Robert, 2008). Thus, developing oral language through conversation, the foundation for decontextualized language, begins when a child is a toddler and can be nurtured at school and at home.

In conclusion, the research findings from this study highlight the importance of EL2 oral language competence for Hispanic ELLs and instructional strategies that enhance English literacy development. My recommendations, based on these findings, support a quadrant approach to address critical issues identified in the research literature on ELLs and English literacy development. Finally, state-mandated high-stakes assessments of minimum skills, like the ETAKS, are one of many challenges facing Hispanic ELLs, educational institutions, and the policy makers who establish language policies and subsequently designate resources to support those policies.

CHAPTER V

CONCLUSION

My dissertation consisted of two studies. Study one, employing a qualitative analytical approach, evaluated the methodological quality of current research on ELLs in two areas: (a) high-stakes testing and (b) oral language as a component skill of reading performance. Study two employed a quantitative approach and examined: (a) the association between oral language and reading performance, and (b) the impact of instructional program model on oral language development.

Findings from the systematic literature reviews (i.e., study one), high-stakes testing and ELLs (presented in Chapter II) and oral language and ELLs (presented in Chapter III) added to the research literature in three ways. First, to date no systematic, replicable review of the literature has evaluated the overall methodological quality of empirical studies of ELLs on the topics of high-stakes testing and oral language. Second, the high-stakes testing and ELLs review appears to be the first study of the impact of high-stakes large-scale achievement testing on ELLs since the passage of NCLB (2001-2009). In previous reviews on the topic of high-stakes testing either ELLs were not included or when included, the review was a narrative, comprehensive study over a period exceeding 20 years. Third, this study is the only systematic review to my knowledge using the same inclusion criteria as the Genesse et al. (2006) synthesis of the research literature on educating ELLs (1980-2003) for the purpose of evaluating the

methodological quality of empirical studies on oral language and ELLs published during the period 2004-2009.

Findings from the review of 11 empirical studies (*n*=3 quantitative, *n*=1 mixed methods, *n*=7 quantitative) on ELLs and high-states testing revealed three meta-themes derived from the research questions: state-wide single assessment accountability, English L2 literacy models and trajectories, and narrowing of curriculum and pedagogy. Issues explored under these meta-themes consisted of the following: student retentions resulting from accountability policies, student perceptions of agency in high-stakes testing environments, language of instruction time allotments in bilingual classrooms, instructional strategies for student achievement, test validity evidence and linguistic bias for ELLs, and teaching to the test. Given the small sample and the number of meta-themes and issues explored, the evidence presented a fragmented picture of ELLs' academic achievement, performance, and outcomes, which in turn affected trend identification in the findings. Therefore, trends were not reported.

Future studies on ELLs and high-stakes testing should explore testing accommodations for ELLs taking high-stakes tests in English and the impact of test accommodations on test validity evidence. In addition, given the availability of student data warehoused by state agencies, researchers should consider using secondary data sets, which are usually large enough for analyzing multiple factors in nested data, (e.g., classrooms, schools). Finally, little evidence is available regarding interventions for ELL struggling readers and high-stakes testing outcomes. Future studies should also explore factors that influence successful outcomes for this group of ELL students.

Findings from the review of 23 empirical studies (*n*=21 quantitative, *n*=2 mixed methods) on ELLs and oral language, revealed two emerging trends in research questions posed by the studies. The first trend was studies examining primary (i.e., L1) language skills as predictors of secondary (i.e., L2) literacy development. In studies of crosslinguistic transference, primary language skill transference was strongest in the early primary grades where ELLs' primary language was the language of reading instruction. The second trend was studies evaluating the associative relationship between literacy component skills and reading comprehension. Findings generally supported a strong association between L2 oral language and L2 reading comprehension. The association appeared weaker in the earlier primary grades than in the upper elementary grades. However, findings from two studies that examined EL2 reading models for ELLs in upper elementary did not agree on the strength of association between oral language and reading comprehension when compared to other reading skills as predictors. Future research should include studies of ELL secondary students, particularly Hispanic ELLs whose dropout rates while declining during the period from 1980-2007 continue to be higher than Whites and Blacks (NCES, 2009). Additionally, few studies have examined oral language in classroom discourse and the development of academic oral language. Both topics represent critical issues in second language acquisition research.

Generally high-stakes and oral language study samples consisted of primary students. Given the paucity of studies of secondary students, future studies should focus on this population. Also noticeably absent were samples that included dual language programs.

The findings from study two, a quantitative study on the relationship between English oral language and ETAKS achievement outcomes for third grade Hispanic ELLs, were presented in Chapter IV. The study used secondary (i.e., archival) longitudinal data collected by a government funded project employing a random sampling method to evaluate instructional programs. Student participants were enrolled in one of four study conditions: SEI-E, SEI-T, TBE-E, and TBE-T.

Oral language was operationalized using vocabulary and listening comprehension measures, whereas reading achievement was operationalized using the ETAKS reading test. Data consisted of 5 time points from kindergarten through the completion of third grade with a latent growth model approach employed for statistical analysis. Results indicated that oral language, especially initial oral language proficiency levels, was a strong predictor of ETAKS reading outcomes. A second strong predictor was oral language trajectory, (i.e., English oral language acquisition rates over the 4-year period). The associative relationship between oral language and reading comprehension performance is supported by evidence from previous studies (Proctor et al., 2005; Snow, 1991; Tong et al., 2008).

Overall results indicated that bilingual education instructional models (experimental and control) experienced the most growth in English verbal and listening comprehension skills compared to the growth rate of the SEI-T group. Additionally students in SEI-E also experienced greater increases in listening comprehension than the SEI-T group. Model results indicated that English oral language accounted for approximately 22-30% of the variability in ETAKS reading scores.

The operationalization of oral language as verbal knowledge and listening comprehension was noted as a limitation in this study. Therefore, this operational definition may limit the generalizability of my study. For future studies, researchers may consider operationalizing oral language employing alternative assessments of *authentic* language tasks as proposed by Bachman (2001; e.g. language samples during classroom discourse).

Finally, two major trends were discussed in this dissertation, each having implications for researchers, policymakers, and practitioners. The following includes lessons learned from each trend. *Trend one*, educational reform driven by standards based assessment, lessons learned were:

- Two analytical techniques–qualitative and quantitative–generate better evidence.
- Systematic literature reviews facilitate critical evaluation and dissemination of research on ELLs to stakeholders by: identifying gaps in the literature, and evaluating the overall methodological quality of the reviewed studies.
- SEM, an analytical technique for more rigorous science, advances theory building and understanding of relationships in complex models.

Trend two, the changing demographics in US public schools, lessons learned were:

- Reading literacy models for ELLs: verbal knowledge and listening comprehension should be included as predictors of English reading

performance, and ELLs' L1 continues to be an important factor in English acquisition.

- Educational policies: funding high quality studies on ELLs, warehousing and making available secondary data from high quality studies, and facilitating research exchanges between researchers and practitioners should be priorities in promoting evidence-based practice.

In conclusion, reading literacy and achievement disparities addressed by NCLB are national concerns. When U.S. school-age targeted subpopulations, including ELLs, do not attain levels of literacy presumed by the global and technology-driven society in which they live, their underdeveloped literacy skills will foreshadow their own success and well-being as citizens. In addition, literacy disparities among their citizens can have profound consequences for those societies that perpetuate them. In closing, a final thought on literacy from Bialystok (2006, p. 107), "… literacy admits children to educational opportunities that shape their futures, a future that enters the genetic code because it moulds the expectancies and outcomes of subsequent generations."

REFERENCES

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

August, D. (2003). *Supporting the development of English literacy in English language learners: Key issues and promising practices* (Report No. 61). Baltimore, MD: John Hopkins University, Center for Research on the Education of Students Placed At Risk. Retrieved January 10, 2009 from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/e5/65.pdf

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice, 20*(1), 50-57.

August, D., Francis, D. J., Hsu, H. A., & Snow, C. (2006a). Assessing reading comprehension in bilinguals. *The Elementary School Journal, 107*(2), 221-238.

August, D., Snow, C., Carlo, M., Proctor, C. P., Rolla de San Francisco, A., Duursma, E., et al. (2006b). Literacy development in elementary school second-language learners. *Topics Language Disorders, 26*(4), 351-364.

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (2001). Some construct validity issues in interpreting scores from performance assessments of language ability. *New Perspectives and Issues in Educational Language Policy, 2001*, 63-90.

Baker, C. (2001). *Foundations of bilingual education and bilingualism* (3rd ed.). Clevedon, UK: Multilingual Matters.

Ballantyne, K.G., Sanderman, A.R., D'Emilio, T. & McLaughlin, N. (2008). *Dual language learners in the early years: Getting ready to succeed in school.* Washington, DC: National Clearinghouse for English Language Acquisition. Retrieved January 11, 2009 from http://www.ncela.gwu.edu/files/uploads/3/DLLs_in_the_Early_Years.pdf

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education* (Vol. 7). Washington, DC: National Association for the Education of Young Children.

Betts, J., Bolt, S., Decker, D., Muyskens, P., & Marston, D. (2009). Examining the role of time and language type in reading development for English language learners. *Journal of School Psychology, 47*(3), 143-166.

Bialystok, E. (2006). Bilingualism at school: Effect on the acquisition of literacy. In P. McCardle & E. Hoff (Eds.), *Childhood bilingualism: Research on infancy through school age* (pp. 107-124). Clevedon, UK: Multilingual Matters.

Booher-Jennings, J. (2008). Learning to label: Socialisation, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education, 29*(2), 149-160.

Bruner, J. S. (1985). Vygotsky: A historical and conceptual perspective. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives* (pp. 21-34). New York: Cambridge University Press.

Calderón, M. (2006). Quality instruction in reading for English language learners. In K. Téllez & H. C. Waxman (Eds.), *Preparing quality educators for English language learners: Research, policies, and practices* (pp. 121-144). Mahwah, NJ: Erlbaum.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics, 1*, 1-47.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J. S., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute. Retrieved January 11, 2009 from http://www.urban.org/UploadedPDF/311230_new_demography.pdf

Cárdenas-Hagan, E., Carlson, C., & Pollard-Durodola, S. (2007). The cross-linguistic transfer of early literacy skills: The role of initial L1 and L2 skills and language of instruction. *Language, Speech, and Hearing Services in Schools, 38*(3), 249-259.

Carlo, M., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D. N., et al. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188-215.

Carver, R. P., & David, A. H. (2001). Investigating reading achievement using a causal model. *Scientific Studies of Reading, 5*(2), 107-140.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1972). *Language and mind*. New York: Harcourt Brace Jovanovich.

Chomsky, N. (1998). *On language*. New York: The New Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, UK: Cambridge University Press.

Cirino, P., Pollard-Durodola, S., Foorman, B., Carlson, C., & Francis, D. (2007). Teacher characteristics, classroom instruction, and student literacy and language outcomes in bilingual kindergartners. *The Elementary School Journal, 107*(4), 341-364.

Cloud, N., Genesse, F., & Hamayan, E. (2000). *Dual language instruction: A handbook for enriched education*. Boston: Heinle & Heinle.

Cummins, J. (1987). Bilingualism, language proficiency, and metalinguistic development. In P. Homel, M. Palikj & D. Aaronson (Eds.), *Childhood bilingualism: Aspects of linguistic, cognitive, and social development* (pp. 57-73). Hillsdale, NJ: Erlbaum.

Cummins, J. (1999). Alternative paradigms in bilingual education research: Does theory have a place? *Educational Researcher, 28*(7), 26-32.

Cummins, J. (2003a). Bilingual education: Basic principles. In J.-M. Dewaele, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles* (pp. 56-66). Clevedon, UK: Multilingual Matters.

Cummins, J. (2003b). Challenging the construction of difference as deficit: Where are identify, intellect, imagination, and power in the new regime of truth? In P. P. Trifonas (Ed.), *Pedagogies of difference: Rethinking education for social change* (pp. 41-60). New York: RoutledgeFalmer.

Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D., et al. (2007). The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics, 28*, 171-190.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.

Escamilla, K., Mahon, E., Riley-Bernal, H., & Rutledge, D. (2003). High-stakes testing, Latinos, and English language learners: Lessons from Colorado. *Bilingual Research Journal, 27*(1), 25-49.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling, 12*(3), 343-367.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavior Research, 42*(3), 504-529.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6*(1), 56-83.

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-312). Greenwich, CT: Information Age.

Fry, R., & Passel, J. (2009). *Latino Children: A majority are U.S.-Born offspring of immigrants*. Washington, DC: Pew Hispanic Center. Retrieved June, 12, 2009, from http://pewhispanic.org/files/reports/110.pdf

García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26*(4), 371-392.

García, E. E., Jensen, B., & Cuéllar, D. (2006). Early academic achievement of Hispanics in the United States: Implications for teacher preparation. *The New Educator, 2*, 123-147.

García, E. E., & Jensen, B. (2007). *Language development and early education of young Hispanic children in the United States*. Tempe, AZ: National Task Force on Early Childhood Education for Hispanics. Retrieved on September 2, 2007, from http://www.ecehispanic.org/work/lang_dev_6August2007.pdf

García, G. E., McKoon, G., & August, D. (2006). Synthesis: Language and literacy assessment. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 583-596). Mahwah, NJ: Lawrence Erlbaum.

Garrard, J. (2007). *Health sciences literature review made easy: The matrix method* (2nd ed.). Boston: Jones and Bartlett.

Genesee, F. (2003). Rethinking bilingual acquisition. In J.-M. Dewaele, A. Housen & L. Wei (Eds.), *Bilingualism: Beyond basic principles* (pp. 204-228). Clevedon, UK: Multilingual Matters.

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

Geva, E., & Zadeh, Z. Y. (2006). Reading efficiency in native English-speaking and English-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading, 10*(1), 31-57.

Glick, J. (1987). Bilingualism: Cognitive and social aspects. In P. Homel, M. Palij & D. Aaronson (Eds.), *Childhood bilingualism: Aspects of linguistic, cognitive, and social development* (pp. 171-179). Hillsdale, NJ: Lawrence Erlbaum.

Goodson, P., Buhi, E., & Dunsmore, S. (2006). Self-esteem and adolescent sexual behaviors, attitudes, and intentions: A systematic review. *Journal of Adolescent Health, 38*(3), 310-319.

Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology, 101*(2), 330-344.

Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review, 28*(1), 135-142.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling*. Greenwich, CT: Information Age Publishing.

Henson, R. K. (1999). Multivariate Normality: What is it and how is it assessed? In B. Thompson (Ed.), *Advances in Social Science Methodology* (Vol. 5, pp. 193-211). Stamford, CT: JAI.

Honoré, H. H. (2008). *Perception of genetic risk in sexual and reproductive decision-making (PGRID) by college students.* Doctoral dissertation, Texas A&M University, College Station.

Hornberger, N. H. (2004). The continua of biliteracy and the bilingual educator: Educational linguistics in practice. *International Journal of Bilingual Education and Bilingualism*, 7(2), 155-171.

Hornberger, N. H. (2008). Continua of biliteracy. In *Encyclopedia of language and education,* (Vol. 9, pp. 275-290). New York: Springer.

Hornberger, N. H., & Johnson, D. C. (2007). Slicing the onion ethnographically: Layers and spaces in multilingual language education policy and practice. *TESOL Quarterly, 41*(3), 509-532.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hult, F. M. (2008). Language ecology and linguistic landscape analysis. In E. Shohamy & D. Gorter (Eds.), *Linguistic landscape: Expanding the scenery* (pp. 88-104). New York: Routledge.

Irby, B. J., Tong, F., Lara-Alecio, R., Meyer, D. J., & Rodriguez, L. (2007). The critical nature of language of instruction compared to observed practices and high-stakes tests in transitional bilingual classrooms. *Research in the Schools, 14*(2), 27-36.

Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics, 30*, 153-185.

John-Steiner, V., & Souberman, E. (1978). Afterword. In M. Cole, S. Scribner, V. John-Steiner & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes* (pp. 121-133). Cambridge, MA: Harvard University Press.

Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly, 35*(4), 524-546.

Kelly, P., Gómez-Bellengé, F., Chen, J., & Schulz, M. (2008). Learner outcomes for English language learner low readers in an early intervention. *TESOL Quarterly, 42*(2), 235-260.

Kieffer, M. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learning in the United States. *Journal of Educational Psychology, 100*(4), 850-868.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* New York: The Guilford Press.

Knight, S. L., & Wiseman, D. L. (2006). Lessons learned from a research synthesis on the effects of teachers' professional development on culturally diverse students. In K. Téllez & H. Waxman (Eds.), *Preparing quality educators for English language learners* (pp. 71-98). Mahwah, NJ: Erlbaum.

Laija-Rodríguez, W., Ochoa, S., & Parker, R. (2006). The crosslinguistic role of cognitive academic language proficiency on reading growth in Spanish and English. *Bilingual Research Journal, 30*(1), 87-106.

Lambert, W. E. (1990). Persistent issues in bilingualism. In B. Harley, P. Allen, J. Cummins & M. Swain (Eds.), *The development of second language proficiency* (pp. 201-218). New York: Cambridge University Press.

Lara-Alecio, R. & Parker, R. (1994). A pedagogical model for transitional English bilingual classrooms, *Bilingual Research Journal, 18* (3&4), 119-133.

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 121(3), 1029-1061.

Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence form cross-state causal-comparative and correlational studies. *Review of Educational Research, 78*(3), 608-644.

Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., & Secada, W. G. (2008). Science achievement of English language learners in urban elementary schools: Results of a first-year professional development intervention. *Journal of Research in Science Teaching, 45*(1), 31-52.

Lesaux, N., & Geva, E. (2006). Synthesis: Development of literacy in language-minority students. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 53-122). Mahwah NJ: Lawrence Erlbaum.

Lindfors, J. W. (1980). *Children's language and learning*. Englewood Cliffs, NJ: Prentice-Hall.

Lindfors, J. W. (2008). *Children's language: Connecting reading, writing, and talk*. New York: Teachers College Press.

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis* (4 ed.). Mahwah, NJ: Erlbaum.

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

Mahon, E. A. (2006). High-stakes testing and English language learners: Questions of validity. *Bilingual Research Journal, 30*(2), 479-497.

Manis, F., Lindsey, K., & Bailey, C. (2004). Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice, 19*(4), 214-224.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-343.

McNeil, L. M., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives, 16*(3), 1-48.

Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal, 30*(2), 521-546.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*(1), 30-43.

Muthén, L.K., & Muthén, B.O. (1998-2007). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Nagy, W. E., García, G. E., Durgunoğlu, A. Y., & Hacin-Bhatt, B. (1993). Spanish-English bilingual students' use of cognates in English reading. *Journal of Reading Behavior, 25*(3), 241-259.

Nakamoto, J., Lindsey, K., & Manis, F. (2008). A cross-linguistic investigation of English language learners' reading comprehension in English and Spanish. *Scientific Studies of Reading, 12*(4), 351-371.

National Assessment of Educational Progress. (2007). *The nation's report card: Reading 2007*. Washington, DC: National Center for Education Statistics. Retrieved November 27, 2008, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496

National Assessment of Educational Progress. (n.d.). *NAEP Inclusion policy*. Washington, DC: National Center for Education Statistics. Retrieved November 2, 2009, from http://nces.ed.gov/nationsreportcard/about/inclusion.asp

National Center for Education Evaluation and Regional Assistance. *Evaluation Studies of the National Center for Education Evaluation and Regional Assistance*. Washington, DC: Institute of Educational Sciences. Retrieved September 12, 2008, from http://ies.ed.gov/ncee/projects/evaluation/ell_projectella.asp

National Center for Education Statistics (2006). *Public elementary and secondary students, staff, schools, and school districts: School year 2003-2004* (NCES 2006-307). Washington, DC: Author. Retrieved July 3, 2009, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006307

National Center for Education Statistics (2009). *The condition of education 2009* (NCES 2009-081). Washington, DC: Author. Retrieved July 25, 2009, from http://nces.ed.gov/pubs2009/2009081.pdf

No Child Left Behind Act of 2001, 20 U.S.C. §1.

Office of English Language Acquisition, Language Enhancement, and Academic Achievement of Limited English Proficient Students (2008). *The biennial report to Congress on the implementation of the Title III State Formula Grant Program: School years 2004-2006* (NCES 2009-081). Washington, DC: Author. Retrieved July 25, 2009, from http:// www.ed.gov/about/offices/list/oela/ title3biennial0406.pdf

Olson, K. (2007). Lost opportunities to learn: The effects of education policy on primary language instruction for English learners. *Linguistics and Education, 18*(2), 121-141.

Ordóñez, C. L., Carlo, M., Snow, C., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology, 94*(4), 719-728.

Ovando, C., Combs, M., & Collier, V. (2006). *Bilingual & ESL classrooms: Teaching in multicultural contexts* (4th ed.). New York: McGraw Hill.

Padrón, Y., & Waxman, H. C. (1999). Effective instructional practices for English language learners. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practices and research* (pp. 171-203). Berkeley, CA: McCutchan.

Pajoohesh, P. (2007). A probe into lexical depth: What is the direction of transfer for L1 literacy and L2 development? *Heritage Language Journal, 5*(1), 117-146.

Popham, W. J. (2008). *The role of assessment in federal education programs*. Washington, DC: Center on Education Policy. Retrieved January 25, 2009, from

http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/43/02/74.pdf

Proctor, C.P., August, D., Carlo, M., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*(1), 159-169.

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246-256.

Roberts, T. A. (2008). Home storybook reading in primary or second language with preschool children: Evidence of equal effectiveness for second-language vocabulary acquisition. *Reading Research Quarterly, 43*(2), 103-130.

Rolla San Francisco, A., Carlo, M., August, D., & Snow, C. E. (2006). The role of language of instruction and vocabulary in the English phonological awareness of Spanish-English bilingual children. *Applied Psycholinguistics, 27*(2006), 229-249.

Saunders, W., Foorman, B., & Carlson, C. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *The Elementary School Journal, 107*(2), 181-198.

Saunders, W. M., & O'Brien, G. (2006). Oral language. In F. Genesee, K. Lindholm-Leary, W. M. Saunders & D. Christian (Eds.), *Educating English Language Learners: A synthesis of research evidence* (pp. 245). New York: Cambridge University

Saville-Troike, M. (1989). *The ethnography of communication: An introduction* (2nd ed.). Oxford, UK: Blackwell.

Scheurich, J. J., & Skrla, L. (2001). Continuing the conversation on equity and accountability: Listening appreciatively, responding responsibly. *Phi Delta Kappan, 83*(4), 322-326.

Scheurich, J. J., Skrla, L., & Johnson, J. E. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan, 82*(4), 293-299.

Shyyan, V., Thurlow, M. L., & Liu, K. K. (2008). Instructional strategies for improving achievement in reading, mathematics, and science for English language learners with disabilities. *Assessment for Effective Intervention, 33*(3), 145-155.

Silverman, R. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *The Elementary School Journal, 107*(4), 365-383.

Singer, J. E., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Skutnabb-Kangas, T., & Mc Carty, T. (2008). Key concepts in bilingual education: Ideological, historical, epistemological, and empirical foundations. In J. Cummins & N. Hornberger (Eds.), *Encyclopedia of language and education,* (Vol. 5, pp. 3-17). New York: Springer.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11.

Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 9-18.

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

Snow, C. E. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review, 53*(2), 165-189.

Snow, C. E. (1991). The theoretical basis for relationship between language and literacy in development. *Journal of Research in Childhood Education, 6*(1), 5-10.

Solórzano, R. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research, 78*(2), 260-329.

Spycher, P. (2009). Learning academic language through science in two linguistically diverse kindergarten classes. *The Elementary School Journal, 109*(4), 359-379.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

Storch, S. A, & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 937-947.

Suen, H. K., & Yu, L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review, 50*(1), 46-65.

Texas Education Agency (1998). *Texas essential knowledge and skills for Spanish language arts and English as a second language*. Austin, TX: Author. Retrieved

September 12, 2008, from http://www.tea.state.tx.us/rules/tac/chapter128/ch128a.html

Texas Education Agency. (2006). *Technical digest 2006-2007*. Austin, TX: Author. Retrieved November 30, 2008, from http://www.tea.state.tx.us/student.assessment/resources/techdig07

Texas Education Agency (2007). *LPAC decision-making process for the Texas assessment program: A procedural manual for the 2007-2008 school year*. Austin, TX: Author. Retrieved September 12, 2008, from http://www.tea.state.tx.us/student.assessment/resources/guides/lpac/LPAC_07_08_section1.pdf

Texas Education Agency. (2008a). *PEIMS core products*. Austin, TX: Author. Retrieved December 30, 2008, from http://www.tea.state.tx.us/adhocrpt/Core_Products/Core_Products.html

Texas Education Agency. (2008b). *2008 comprehensive annual report on Texas public schools* (Document No. GE09 601 03). Austin, TX: Author. Retrieved December 5, 2008, from http://www.tea.state.tx.us/research/pdfs/2008_comp_annual.pdf

Texas Education Agency. (2008c). *Grade-level retention in Texas public schools, 2006-07* (Document No. GE09 601 01). Austin, TX: Author. Retrieved November 30, 2008, from http://www.tea.state.tx.us/research/pdfs/retention_2006-07.pdf

Texas Education Agency. (2008d). *Graduation requirements for Texas students*. Austin TX: Author. Retrieved October 15, 2009, from

http://ritter.tea.state.tx.us/student.assessment/resources/grad/2008-09_GradBro.pdf

Texas Education Agency. (2008e). *2007-2008 Academic excellence indicator system*. Austin TX: Author. Retrieved February 28, 2009, from http://ritter.tea.state.tx.us/perfreport/aeis/2008/district.srch.html

Texas Education Agency. (2009a). *Grade Placement Committee Manual 2009-20010*. Austin, TX: Author. Retrieved November 2, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/ssi/GPCManual.pdf

Texas Education Agency. (2009b). *Enrollment in Texas public schools 2007-2008*. Austin, TX: Author. Retrieved February 28, 2009, from http://ritter.tea.state.tx.us/research/pdfs/enrollment_2007-08.pdf

Thompson, B. (1990). MULTINOR: A FORTRAN program that assists in evaluating multivariate normality. *Educational and Psychological Measurement, 50*, 845-848.

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261-283). Washington, DC: American Psychological Association.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford.

Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., & Kwok, O. (2008). Accelerating early academic oral English development in transitional bilingual and structured

English immersion programs. *American Educational Research Journal, 45*(4), 1011-1044.

Tsang, S., Katz, A., & Stack, J. (2008). Achievement testing for English language learners, ready or not? *Education Policy Analysis Archives, 16*(1), 1-29.

Valenzuela, A. (2005). Introduction: The accountability debate in Texas: Continuing the conversation. In A. Valenzuela (Ed.), *Leaving children behind: How "Texas-style" accountability fails Latino youth* (pp. 1-32). Albany, NY: State University of New York Press.

Vaughn, S., Cirino, P., Linan-Thompson, S., Mathes, P., Carlson, C., Cardenas-Hagan, E., et al. (2006b). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal, 43*(3), 449-487.

Vaughn, S., Linan-Thompson, S., Mathes, P., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities, 39*(1), 56-73.

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*(2), 153-180.

Verhoeven, L. (1997). Acquisition of literacy by immigrant children. In C. Pontecorvo (Ed.), *Writing development: An interdisciplinary view* (pp. 219-240). Amsterdam: Benjamins.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4*(4), 313-330.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wall, D., & Horák, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education: Principles, Policy & Practice, 14*(1), 99-116.

Wong Fillmore, L., & Snow, C. (2002). What teachers need to know about language. In C. T. Adger, C. Snow & D. Christian (Eds.), *What teachers need to know about language* (pp. 7-54). McHenry, IL: Center for Applied Linguistics and Delta Systems.

Woodcock, R. W. (1991a). *Woodcock Language Proficiency Battery-Revised, English Form.* Itasca, IL: Riverside Publishing.

Woodcock, R. W. (1991b). *Woodcock Language Proficiency Battery-Revised Examiner's Manual.* Itasca, IL: Riverside Publishing.

APPENDIX A

ENGLISH LANGUAGE LEARNERS, HIGH-STAKES TESTING, AND ACCOUNTABILITY IN TEXAS

The Texas Assessment of Knowledge and Skills

For Texas public school students, the third grade represents a milestone in their schooling, their first administration of the Texas Assessment of Knowledge and Skills (TAKS) test. The TAKS, a criterion-referenced assessment aligned to state curriculum standards, is an annual measure of student progress (Texas Education Agency, 2008b). The current Texas accountability system mandates annual state assessments of reading and math achievement beginning at third grade and continuing through the eleventh grade.

Grade promotion in Texas is based on student achievement, which for grades 3-11 under the original Student Success Initiative (SSI) grade advancement rules was measured by annual state assessments like the TAKS test (Texas Education Agency, 2006). For seven years, from spring 2003 through spring 2009, third graders were required to pass the TAKS reading test in order to advance to the fourth grade. Under the revised SSI grade advancement rules, beginning in the school year 2009-2010, only 5th and 8th graders must pass the TAKS reading and math tests to advance to the next grade (Texas Education Agency, 2009a). Students who fail the TAKS and whose TAKS performance is a grade advancement requirement can only advance to the next grade by

unanimous vote of the grade placement committee members at their respective schools (Texas Education Agency, 2008c).

In addition, state-wide assessments, called exit level academic tests, are a graduation requirement for high school students. In the 11th grade, students must pass exit level academic tests in English language arts, mathematics, science, and social studies (TEA, 2008d). English Language Learners (ELL), are not exempt from these high school graduation requirements.

Thus, the cornerstone of the Texas accountability system in schools, the Academic Excellence Indicator System (AEIS), albeit controversial (McNeil, Coppola, Radigan, & Heilig, 2008; Scheurich & Skrla, 2001; Scheurich, Skrla, & Johnson, 2000; Valenzuela, 2005), is student performance on state-mandated large-scale assessments with the school as the unit and focal point of analysis.

English Language Learners and the TAKS

For third grade Hispanic ELLs, program placement and language of instruction determine the language version of the TAKS test administered to students. In Texas school districts, the Language Proficiency Assessment Committee (LPAC)[12] makes decisions about TAKS test exemptions and TAKS test language version (i.e., English or Spanish) for all limited English proficient (LEP) [13] or ELL students. The Spanish TAKS is only available from grades 3 through 6 (TEA, 2007). All ELL students placed in

[12] The LPAC can exempt immigrant LEP students for three years, from the date of their entry into the United States. LEP students are postponed but not exempted from the exit level TAKS tests (Texas Administration Code §101.1005).

[13]LEP is used alternatively with English Language Learners. The term LEP is used in the Texas Education Code (TEC).

English as a second language (ESL) program models must take the TAKS test in English. The TAKS test scores of ELLs are reported under the category entitled *LEP*, one of the subgroups whose performance scores are disaggregated in state reporting (i.e., AEIS) and calculated in the state formula to determine school and district ratings.

National Assessment of Educational Progress and Texas ELLs

The National Assessment of Educational Progress (NAEP, 2007) reported that on NAEP fourth grade reading tests 62% of Texas ELLs scored below the *basic* level of reading achievement while 38% of Texas ELLs scored at the *basic* level of reading achievement.[14] NAEP fourth grade reading scores for all U.S. ELLs were 70% (below *basic* level) and 30% (at *basic* level). On the eighth grade NAEP reading test, 80% of Texas ELLs scored below the *basic* level of reading achievement, and 20% of Texas ELLs scored at the *basic* level. Nationally, NAEP eighth reading scores for all ELLs were 71% (below *basic* level) and 29% (at *basic* level). Therefore, 100% of all ELLs' reading performance on the NAEP reading tests was below the *proficient* level for mastery of grade level knowledge and skills in reading.

Grade Retention and ELLs

Finally, the Texas Education Agency (TEA) reported that for the years 1994-2007 ELL retention rates were higher than non-ELL retention rates in all the elementary

[14] NAEP employs three descriptor categories for grade level achievement: basic, proficient, and advanced. Basic denotes partial mastery of grade level achievement standards, i.e. knowledge and skills, in the content area of the NAEP test (e.g. math, reading). NAEP reading tests are in English with some accommodations allowed for ELLs. Some of the accommodations permitted by NAEP are the following: reading the test instructions in the native language, small group administration , and extended time to name a few. However, no bilingual dictionaries, translation of test items to ELLs' native language, or bilingual versions (Spanish/English) are allowed (NAEP Inclusion Policy, n.d.)

grades except kindergarten (TEA, 2008c, 2009b). By the year 2006-2007, the ELL retention rate in grades K-6 was 85% higher than the non-ELL retention rate (4.6% for ELLs compared to 2.7% for non-ELLs). Thus, ELL retention rates and English reading achievement are critical issues for ELLs and their families and for Texas policymakers, district and school administrators, and teacher practitioners. Consequently, identifying foundation skills that predict reading achievement and program models that support English reading literacy is imperative. Oral language appears to be one of those foundation skills that correlate to reading achievement (Lesaux & Geva, 2006).

APPENDIX B

INSTITUTIONAL REVIEW BOARD DOCUMENTS

TEXAS A&M UNIVERSITY
DIVISION OF RESEARCH AND GRADUATE STUDIES - OFFICE OF RESEARCH COMPLIANCE

1186 TAMU, General Services Complex
College Station, TX 77843-1186
750 Agronomy Road, #3500

979.458.1467
FAX 979.862.3176
http://researchcompliance.tamu.edu

Institutional Biosafety Committee | Institutional Animal Care and Use Committee | Institutional Review Board

DATE: 23-Sep-2008

MEMORANDUM

TO: ACOSTA, SANDRA

FROM: Office of Research Compliance
Institutional Review Board

SUBJECT: Initial Review

Protocol Number:	2008-0540
Title:	High Stakes Reading Assessment and English Oral Language Development: A Longitudinal Study of Third Grade English Language Learners in Texas
Review Category:	Expedited
Approval Period:	23-Sep-2008 **To** 22-Sep-2009

Approval determination was based on the following Code of Federal Regulations:

(5) Research involving materials (data, documents, records, or specimens) that have been collected or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis).

(Note: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(4). This listing refers only to research that is not exempt.)

Provisions: This is a study of exisiting data from Protocol 2003-0414.

This research project has been approved for one (1) year. As principal investigator, you assume the following responsibilities

1. **Continuing Review:** The protocol must be renewed each year in order to continue with the research project. A Continuing Review along with required documents must be submitted 30 days before the end of the approval period. Failure to do so may result in processing delays and/or non-renewal.
2. **Completion Report:** Upon completion of the research project (including data analysis and final written papers), a Completion Report must be submitted to the IRB Office.
3. **Adverse Events:** Adverse events must be reported to the IRB Office immediately.

4. **Amendments:** Changes to the protocol must be requested by submitting an Amendment to the IRB Office for review. The Amendment must be approved by the IRB before being implemented.
5. **Informed Consent:** Information must be presented to enable persons to voluntarily decide whether or not to participate in the research project.

This electronic document provides notification of the review results by the Institutional Review Board.

APPENDIX C-1

ALPHABETIZED LIST OF ENGLISH LANGUAGE LEARNERS AND HIGH-STAKES ABSTRACTED STUDIES

1. Betts, J., Bolt, S., Decker, D., Muyskens, P., & Marston, D. (2009). Examining the role of time and language type in reading development for English language learners. *Journal of School Psychology, 47*(3), 143-166.

2. Booher-Jennings, J. (2008). Learning to label: Socialisation, gender, and the hidden curriculum of high-stakes testing. *British Journal of Sociology of Education, 29*(2), 149-160.

3. Escamilla, K., Mahon, E., Riley-Bernal, H., & Rutledge, D. (2003). High-stakes testing, Latinos, and English language learners: Lessons from Colorado. *Bilingual Research Journal, 27*(1), 25-49.

4. Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review, 28*(1), 135-142.

5. Irby, B. J., Tong, F., Lara-Alecio, R., Meyer, D. J., & Rodriguez, L. (2007). The critical nature of language of instruction compared to observed practices and high-stakes tests in transitional bilingual classrooms. *Research in the Schools, 14*(2), 27-36.

6. Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., & Secada, W. G. (2008). Science achievement of English language learners in urban elementary schools: Results

of a first-year professional development intervention. *Journal of Research in Science Teaching, 45*(1), 31-52.

7. Mahon, E. A. (2006). High-stakes testing and English language learners: Questions of validity. *Bilingual Research Journal, 30*(2), 479-497.

8. Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal, 30*(2), 521-546.

9. Olson, K. (2007). Lost opportunities to learn: The effects of education policy on primary language instruction for English learners. *Linguistics and Education, 18*(2), 121-141.

10. Shyyan, V., Thurlow, M. L., & Liu, K. K. (2008). Instructional strategies for improving achievement in reading, mathematics, and science for English language learners with disabilities. *Assessment for Effective Intervention, 33*(3), 145-155.

11. Tsang, S.-L., Katz, A., & Stack, J. (2008). Achievement testing for English language learners, ready or not? *Education Policy Analysis Archives, 16*(1), 1-29.

APPENDIX C-2

ALPHABETIZED LIST OF ENGLISH LANGUAGE LEARNERS AND ORAL LANGUAGE ABSTRACTED STUDIES

1. Cárdenas-Hagan, E., Carlson, C., & Pollard-Durodola, S. (2007). The cross-linguistic transfer of early literacy skills: The role of initial L1 and L2 skills and language of instruction. *Language, Speech, and Hearing Services in Schools, 38*(3), 249-259.

2. Cirino, P., Pollard-Durodola, S., Foorman, B., Carlson, C., & Francis, D. (2007). Teacher characteristics, classroom instruction, and student literacy and language outcomes in bilingual kindergartners. *The Elementary School Journal, 107*(4), 341-364.

3. Gottardo, A., Collins, P., Baciu, I., & Gebotys, R. (2008). Predictors of grade 2 word reading and vocabulary learning from grade 1 variables in Spanish-speaking children: Similarites and differences. *Learning Disabilities Research & Practice, 23*(1), 11-24.

4. Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology, 101*(2), 330-344.

5. Kelly, P., Gómez-Bellengé, F., Chen, J., & Schulz, M. (2008). Learner outcomes for English language learner low readers in an early intervention. *TESOL Quarterly, 42*(2), 235-260.

6. Kieffer, M. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learning in the United States. *Journal of Educational Psychology, 100*(4), 850-868.

7. Laija-Rodríguez, W., Ochoa, S., & Parker, R. (2006). The crosslinguistic role of cognitive academic language proficiency on reading growth in Spanish and English. *Bilingual Research Journal, 30*(1), 87-106.

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

9. Manis, F., Lindsey, K., & Bailey, C. (2004). Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice, 19*(4), 214-224.

10. Miller, J., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*(1), 30-43.

11. Nakamoto, J., Lindsey, K., & Manis, F. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing, 20*(7), 691-719.

12. Nakamoto, J., Lindsey, K., & Manis, F. (2008). A cross-linguistic investigation of English language learners' reading comprehension in English and Spanish. *Scientific Studies of Reading, 12*(4), 351-371.

13. Proctor, C., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246-256.

14. Proctor, C., August, D., Carlo, M., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*(1), 159-169.

15. Roberts, T. A. (2008). Home storybook reading in primary or second language with preschool children: Evidence of equal effectiveness for second-language vocabulary acquisition. *Reading Research Quarterly, 43*(2), 103-130.

16. Saunders, W., Foorman, B., & Carlson, C. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *The Elementary School Journal, 107*(2), 181-198.

17. Silverman, R. (2007). Vocabulary Development of English-language and English-only learners in kindergarten. *The Elementary School Journal, 107*(4), 365-383.

18. Spycher, P. (2009). Learning academic language through science in two linguistically diverse kindergarten classes. *The Elementary School Journal, 109*(4), 359-379.

19. Tong, F., Irby, B. J., Lara-Alecio, R., & Mathes, P. (2008a). English and Spanish acquisition by Hispanic second graders in developmental bilingual programs: A 3-year longitudinal randomized study. *Hispanic Journal of Behavioral Sciences, 30*(4), 500-529.

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

21. Vaughn, S., Cirino, P., Linan-Thompson, S., Mathes, P., Carlson, C., Cárdenas-Hagan, E., et al. (2006a). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal, 43*(3), 449-487.

22. Vaughn, S., Linan-Thompson, S., Mathes, P., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006b). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities, 39*(1), 56-73.

23 . Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006c). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*(2), 153-180.

APPENDIX D

ALPHABETIZED LIST OF JOURNALS FOR JOURNAL SEARCH

1. American Educational Research Journal
2. Bilingual Research Journal
3. Contemporary Educational Psychology
4. Developmental Psychology
5. Dyslexia
6. Educational Researcher
7. Harvard Educational Review
8. Hispanic Journal of Behavioral Sciences
9. Journal of Educational Psychology
10. Journal of Reading Education
11. Journal of Research in Childhood Education
12. Journal of Research in Reading
13. Journal of Speech, Language, and Hearing Research
14. Language Learning
15. Language, Speech, and hearing Services in Schools
16. Learning Disabilities Research and Practice
17. Literacy
18. Reading Research Quarterly
19. Review of Educational Research

20. Science

21. Scientific Studies of Reading

22. TESOL Quarterly

23. The Elementary School Journal

24. The Modern Language Journal

25. Topics in Language Disorders

APPENDIX E-1

ABSTRACTION SHEET FOR HIGH-STAKES TESTING AND ENGLISH LANGUAGE LEARNERS REVIEWED STUDIES

Study ID # _____
Coder: _____

High-Stakes Reading Assessment and English Oral language Development for Third Grade English Language Learners: A Systematic Review of the Literature

Acosta 2009
Texas A&M University – Dept. Educational Psychology

Abstraction Sheet

Date of Abstraction: _____ Research Topic: High-Stakes Testing and English Language Learners (ELLs)

Bibliographic Information:

Citation:

Research Question(s):	**Journal Article Page no.**

☐ Discusses the relationship between EL2 oral language and EL2 reading test (state-mandated)
☐ Discusses the relationship between EL1 oral language (non-English) and EL1 (non-English) reading test (state-mandated)
☐ Discusses the relationship between oral language/verbal knowledge and other content areas (math/science/social studies, writing)
☐ Discusses the relationship between classroom discourse/instructional conversation and state-mandated tests
☐ Other

Purpose	**Journal Article Page no.**

VARIABLES	**Journal Article Page no.**
Dependent Variable(s)	**Independent Variable(s)**

High-Stakes Test	**Journal Article Page no.**
Test:	State:

Language(s)	**Journal Article Page no.**
ELLs' L1:	Language of Test:

CRITERIA

Ql= Qualitative Studies Qn=Quantitative Studies Wave(s) = time point(s)

Criterion 1 – Operationalization of variable(s): _____ **pts.** **Journal Article Page no.** _____

○ Did NOT provide operational definition of variable(s) for Qn OR construct/phenomenon for Ql; or operational definition implied (0)

○ Provided partial operational definition of variable(s) for Qn OR construct/phenomenon for Ql (1)

○ Explicit operational definition of variable(s) for Qn OR construct/phenomenon for Ql (2)

Terminology: Use of the word operationalize (or derivatives) ○ YES ○ NO

Note: Implicit - cites the research literature for various conceptual definitions of the construct, no linkage to the selection of the variables to be measured; Partial - provides a conceptual definition and the linkage to the variable(s); Explicit: provides a conceptual definition, the linkage to the variable, an explanation of the instrument and its properties to accurately measure the variable.
For qualitative: the researcher must explain how the instrument (the interview or observation)will capture the "complexity" (i.e. the various facets of the construct). The most important part of the operational definition is the description of the instrument's capacity to quantify or describe the characteristic and the degree of its presence.

Reviewer's Notes/Comments:

Criterion 2 – Theory: _____ **pts.** **Journal Article Page no.** _____
(Theory = theory, theoretical framework, conceptual framework)

○ Did not use theory; or theory implied (0)

Implicit = no theory-based explanation or cites empirical literature for relationship between the IV/DV for Qn OR characteristics/factors for Ql

○ Explicit explanation of theory (2)

Explicit = cites theory(s) or conceptual framework for relationship between the IV/DV for Qn OR characteristics/factors for Ql

Terminology: Use of the word theory (or derivatives) ○ YES ○ NO

Note: In Ql research the theoretical or conceptual framework is not a priori as in Qn research but rather posteriori or during the research.

Reviewer's Notes/Comments:

Criterion 3 – Instrumentation: _____ **pts.** **Journal Article Page no.** _____
(Instrumentation = survey, questionnaire, rubric, checklist, tests/assessments, etc.)

○ Used instruments (1)
Included description of instrument; Reliability/validity NOT reported

○ Used instruments (2)
Included description of instrument; Reliability/validity reported (citation or statistic)

***Note:** The human is the instrument in Ql research and should have a minimum of 1 pt. To earn full points procedures to achieve a higher reliability and validity must be addressed.*

Reviewer's Notes/Comments:

Criterion 4 - Research Paradigm: _____ **pts.** **Journal Article Page no.** _____

○ Qualitative (1) ○ Quantitative (1) ○ Mixed Methods (2)

Reviewer's Notes/Comments:

Criterion 5 - Research Design: _____ **pts.** **Journal Article Page no.** _____

○ Idiographic/qualitative (1) ○ Cross-Sectional (1) ○ Pretest-Posttest (2) ○ Longitudinal-change 3 waves or > (3)

Reviewer's Notes/Comments:

Criterion 6 - Sampling: _____ **pts.** **Journal Article Page no.** _____

○ Non-random or convenience sample (1) ○ Non-random/convenience sample + comparison group (2)

○ Random sample/Experimental or quasi experimental (3)

Reviewer's Notes/Comments:

Criterion 7 – Sample Size: _____ **pts.** **Journal Article Page no.** _____

○ <30 = Small (0) ○ <30 = Small (1) Qualitative only ○ >30 and <300 = Medium (1) ○ 300 or >= Large (2)

N (total) =	Ethnicity(ies):	Ages/Grade(s):	Year Data Collected:
	Languages:		

Reviewer's Notes/Comments:
Length of study:

Criterion 8 – Language Groups: _____ **pts.** **Journal Article Page no.** _____

○ English Language Learners ELLs Only (1) ○ EO & ELLs (1) ○ Dual Language (2)

Reviewer's Notes/Comments:

Criterion 9 - Participants: _____ **pts.** **Journal Article Page no.** _____

○ Early childhood, pre-kindergarten through 2nd grades (1)

○ Upper elementary, 4th through 6th grades (2)

○ Primary school, pre-K through 5th or 6th (1)

○ Middle school, 6th through 8th grades (3)

○ High school 9th through 12th grades (3)

Reviewer's Notes/Comments:

Criterion 10 – Data Analysis: _____ **pts.** **Journal Article Page no.** _____

○ Univariate Statistics: ANOVA, descriptive-frequencies and mean distributions; percentages (1)

○ Qualitative: content analysis; emergent themes analysis; grounded theory; etc. (2)

○ Bivariate Statistics: Correlation or Crosstabulation, Chi-Square (2)

○ Multiple/Logistic Regression (3)

○ Multivariate Statistics: Canonical correlatons analysis; discrimination function analysis; path analysis,factor analysis, MANOVA, SEM, LGM, HLM (4)

Reviewer's Notes/Comments:

Methodological Characteristics Scores		
	Characteristic(s)	***Score***
C1	Operationalization of Variables /Phenomenon	
C2	Theory	
C3	Instrumentation	
C4	Research Paradigm	
C5	Research Design	
C6	Sampling	
C7	Sample size (small, medium, large)	
C8	Language Groups (EO, ELL, Dual Language)	
C9	Participants (Early childhood, upper elementary, middle school, high school)	
C10	Data Analysis	
Total Score (Maximum possible score = 25)		

Findings			
Test / Finding / Relationship	DV	ES	Direction of Finding (-1, 0 or +1)

Does conclusion inappropriately imply causality? ○ Yes ○ No ○ Not Applicable

Quote *(include journal article page number)*

Comments/Notes/ Potential Quotes *(for quotes, include journal article page number)*

Themes *(constructs, descriptors not appearing in the criteria or findings)*

APPENDIX E-2

ABSTRACTION SHEET FOR ORAL LANGUAGE AND ENGLISH LANGUAGE LEARNERS REVIEWED STUDIES

Study ID # _____

High-Stakes Reading Assessment and English Oral language Development for Third Grade English Language Learners: A Systematic Review of the Literature

Acosta 2009
Texas A&M University – Dept. Educational Psychology

Abstraction Sheet

Date of Abstraction: _____ Research Topic: Oral Language Development and English Language Learners (ELLs)

Bibliographic Information:

Citation:

Research Question(s):	**Journal Article Page no.**

☐ Discusses the relationship between oral language and reading
☐ Discusses the relationship between EL1 oral language (non-English) and EL1 (non-English) reading test (state-mandated)
☐ Discusses the relationship between verbal knowledge (oral language) and academic/domain language
☐ Discusses the relationship between oral language and classroom discourse/instructional conversation
☐ Discusses the relationship between crosslinguistic transfer and L2 /L1 literacy
☐ Other

Purpose	**Journal Article Page no.**

VARIABLES	**Journal Article Page no.**
Dependent Variable(s)	**Independent Variable(s)**

CRITERIA

Ql= Qualitative Studies Qn=Quantitative Studies Wave(s) = time point(s)

Criterion 1 – Operationalization of variable(s): _____ **pts.** **Journal Article Page no.** _____

○ Did NOT provide operational definition of variable(s) for Qn OR construct/phenomenon for Ql; or operational definition implied (0)

○ Provided partial operational definition of variable(s) for Qn OR construct/phenomenon for Ql (1)

○ Explicit operational definition of variable(s) for Qn OR construct/phenomenon for Ql (2)

Terminology: Use of the word operationalize (or derivatives) ○ YES ○ NO

Note: Implicit - cites the research literature for various conceptual definitions of the construct, no linkage to the selection of the variables to be measured; Partial - provides a conceptual definition and the linkage to the variable(s); Explicit: provides a conceptual definition, the linkage to the variable, an explanation of the instrument and its properties to accurately measure the variable.
For qualitative: the researcher must explain how the instrument (the interview or observation)will capture the "complexity" (i.e. the various facets of the construct). The most important part of the operational definition is the description of the instrument's capacity to quantify or describe the characteristic and the degree of its presence.

Reviewer's Notes/Comments:

Criterion 2 – Theory: _____ **pts.** **Journal Article Page no.** _____
(Theory = theory, theoretical framework, conceptual framework)

○ Did not use theory; or theory implied (0)

Implicit = no theory-based explanation or cites empirical literature for relationship between the IV/DV for Qn OR characteristics/factors for Ql

○ Explicit explanation of theory (2)

Explicit = cites theory(s) or conceptual framework for relationship between the IV/DV for Qn OR characteristics/factors for Ql

Terminology: Use of the word theory (or derivatives) ○ YES ○ NO

Note: In Ql research the theoretical or conceptual framework is not a priori as in Qn research but rather posteriori or during the research.

Reviewer's Notes/Comments:

Criterion 3 – Instrumentation: _____ **pts.** **Journal Article Page no.** _____
(Instrumentation = survey, questionnaire, rubric, checklist, tests/assessments, etc.)

○ Used instruments (1)
Included description of instrument; Reliability/validity NOT reported

○ Used instruments (2)
Included description of instrument; Reliability/validity reported (citation or statistic)

***Note:** The human is the instrument in Ql research and should have a minimum of 1 pt. To earn full points procedures to achieve a higher reliability and validity must be addressed.*

Reviewer's Notes/Comments:

Criterion 4 - Research Paradigm: _____ **pts.** **Journal Article Page no.** _____

○ Qualitative (1) ○ Quantitative (1) ○ Mixed Methods (2)

Reviewer's Notes/Comments:

Criterion 5 - Research Design: _____ **pts.** **Journal Article Page no.** _____

○ Idiographic/qualitative (1) ○ Cross-Sectional (1) ○ Pretest-Posttest (2) ○ Longitudinal-change 3 waves or > (3)

Reviewer's Notes/Comments:

Criterion 6 - Sampling: _____ **pts.** **Journal Article Page no.** _____

○ Non-random or convenience sample (1) ○ Non-random/convenience sample + comparison group (2)

○ Random sample/Experimental or quasi experimental (3)

Reviewer's Notes/Comments:

Criterion 7 – Sample Size: _____ **pts.** **Journal Article Page no.** _____

○ <30 = Small (0) ○ <30 = Small (1) Qualitative only ○ >30 and <300 = Medium (1) ○ 300 or >= Large (2)

N (total) =	Ethnicity(ies):	Ages/Grade(s):	Year Data Collected:
	Languages:		

Reviewer's Notes/Comments:
Length of study:

Criterion 8 – Language Groups: _____ **pts.** **Journal Article Page no.** _____

○ English Language Learners ELLs Only (1) ○ EO & ELLs (1) ○ Dual Language (2)

Reviewer's Notes/Comments:

Criterion 9 - Participants: _____ **pts.** **Journal Article Page no.** _____

○ Early childhood, pre-kindergarten through 2nd grades (1)

○ Upper elementary, 4th through 6th grades (2)

○ Primary school, pre-K through 5th or 6th (1)

○ Middle school, 6th through 8th grades (3)

○ High school 9th through 12th grades (3)

Reviewer's Notes/Comments:

Criterion 10 – Data Analysis: _____ **pts.** **Journal Article Page no.** _____

○ Univariate Statistics: ANOVA, descriptive-frequencies and mean distributions; percentages (1)

○ Qualitative: content analysis; emergent themes analysis; grounded theory; etc. (2)

○ Bivariate Statistics: Correlation or Crosstabulation, Chi-Square (2)

○ Multiple/Logistic Regression (3)

○ Multivariate Statistics: Canonical correlatons analysis; discrimination function analysis; path analysis,factor analysis, MANOVA, SEM, LGM, HLM (4)

Reviewer's Notes/Comments:

Methodological Characteristics Scores		
	Characteristic(s)	***Score***
C1	Operationalization of Variables /Phenomenon	
C2	Theory	
C3	Instrumentation	
C4	Research Paradigm	
C5	Research Design	
C6	Sampling	
C7	Sample size (small, medium, large)	
C8	Language Groups (EO, ELL, Dual Language)	
C9	Participants (Early childhood, upper elementary, middle school, high school)	
C10	Data Analysis	
Total Score (Maximum possible score = 25)		

Findings			
Test / Finding / Relationship	DV	ES	Direction of Finding (-1, 0 or +1)

Does conclusion inappropriately imply causality? ○ Yes ○ No ○ Not Applicable

Quote *(include journal article page number)*

Comments/Notes/ Potential Quotes *(for quotes, include journal article page number)*

Themes *(constructs, descriptors not appearing in the criteria or findings)*

APPENDIX F

OVERVIEW OF THEORIES RELEVANT TO SECOND LANGUAGE ACQUISITION

All children acquire their first language at about the same rate so that by kindergarten children are proficient in their first expressive system,[15] generally the language spoken at home (Lindfors, 1980, 2008). Albeit having successfully acquired their first expressive system, ELLs' second language acquisition timelines and outcomes can be affected by many factors. Thus, the purview of second language acquisition theorists is to incorporate the *facts* of bilingualism into their theories (Genesse, 2003) by describing developmental patterns, identifying associated factors, and explaining models that capture the second language acquisition and literacy process (Ellis, 2008).

Overviews of relevant theories are important for two primary reasons. First, these overviews provide the foundation for understanding current theories or conceptual models (e.g. ELLs' reading performance). Second, they are the progenitors for establishing paradigms that allow future researchers to develop parsimonious and coherent language models for their empirical studies of second language acquisition and literacy development. The three theorists that I will consider in the next section are Chomsky, Vygotsky, and Hornberger, whose theories form the basis for the theoretical assumptions about second language acquisition and learning in my study.

---

[15]The first expressive system, sometimes referred to as oracy, consists of speaking and listening. The second expressive system, literacy, consists of reading and writing. These four basic language abilities: listening, speaking reading and writing are classified into the dimensions of receptive skills (i.e., listening and reading) and expressive skills (i.e., speaking and writing; Baker, 2001).

Chomsky's theories on the biological nature of language define the first theoretical lens of my study. Chomsky made two substantial contributions to second language development theory. First, as the architect of the *innatist* view of language, Chomsky (1972, 1998, 2000) posited a biologically determined language system. He attributed the universal language acquisition timeline and a child's capacity to build linguistic competence in a complex language system to the faculty for language present at birth and shared by all humans and humans alone. Chomsky's Language Acquisition Device (LAD) referred to the group of theories that explain language growth or change from the initial state of the faculty for language (i.e., the Universal Grammar) to later states that incorporated language growth (Chomsky, 2000).

In addition, Chomsky (1965) first introduced the constructs *competence* and *performance* in the context of language development. Competence for Chomsky only referred to the domain of grammar, while performance was described as the interactions between grammar and "a set of nongrammatical psychological factors bearing on language use" (e.g., speech production; Canale & Swain, 1980, p. 3). In second language acquisition theory, language competence and language performance became important concepts in measuring language proficiency levels and bilingualism (Baker, 2001).

Building on Chomsky's earlier conceptualization of competence, years later, Canale and Swain (1980) proposed a model of language competence that included grammar and other extra-linguistic components (e.g., strategic competence, verbal and nonverbal strategies employed to maintain conversation). In 1990 Bachman developed a

second major theory of language competence, which unlike the Canale and Swain model incorporated language performance. In addition to making language competence components interrelated, Bachman also added strategic competence and used it to model the dynamic editing function of language. The metalinguistic activities incorporated under strategic competence were the strategies that an individual employed to plan, execute, and subsequently assess communications (Baker, 2001). Having established these criteria for language competence, Bachman proposed that language performance, also known as alternative performance, could no longer be assessed with only paper and pencil tasks but should be assessed employing *authentic* language tasks. Thus, these language samples (i.e., observations) would capture the dynamic aspects of language competence (Bachman, 2001).

The second theoretical lens is Vygotsky's sociocultural theory of development. Vygotsky conceptualized two levels of cognitive development, elementary and complex cognitive processes. Elementary cognitive processes are biological in origin and characteristic of infants and very young children. Complex cognitive processes have sociocultural origins and enable thoughts, metacognition, and concept formation through the transformation of linguistic and cognitive structures (Berk & Winsler, 1995; Vygotsky, 1962).

According to Vygotsky the primary source of change (i.e., learning) is development impelled by social interaction. The initial portal to cognition for both first and second language learners is oral language which serves as a mediator of communication. Word meanings are negotiated and shaped through these social

interactions. Thus, information becomes knowledge through action and practice guided (i.e., scaffolded) by adults (Glick, 1987; Vygotsky, 1962, 1978). Bruner (1985) proposed that in school settings curriculum, learning, and teaching equate to Vygotsky's *props*, *processes*, and *procedures* for scaffolding cognitive growth and guiding children to independent performance.

For purposes of encapsulating the first two theoretical lenses, Chomsky and Vygotsky asked the quintessential question, "*What is the nature of language development in human beings?*" Together their theories (i.e., Chomsky's theory of the human language faculty and Vygotsky's theory of language development and thought) form a *mega theory* of first language acquisition and its relationship to cognitive development. These classic theories continue to have relevance for second language acquisition and pedagogy as researchers examine the effects of bilingualism on language development in young children.

One issue regarding language development and bilingualism is whether learning two languages can have possible deleterious psychological effects on a child and the child's academic achievement (Genesee, 2003). Generally bilingual children are divided into two groups. *Simultaneous* bilinguals, also referred to as *infant bilingualism*, refers to children who are bilingual or multilingual from birth (e.g. more than one language is spoken in the home or a second language is spoken by the child's primary caregiver). In contrast, *sequential* bilinguals are young children who acquire a second language later (Baker, 2001). An example of sequential bilingualism would be primary school ELLs

enrolled in bilingual education programs who are receiving instruction in two languages, their mother tongue (L1) and a second language (L2).

Chomsky's theory of the human language faculty does not support behaviorists' and English-only advocates' concerns that young ELLs' academic achievement will be compromised when they receive instruction in two languages (Cummins, 2003a, 2003b). Conversely, the human language faculty theory does appear to support emerging research on simultaneous bilingual infants, that "human infants possess the biological capacity to acquire two languages as normally as one" (Genesee, 2003; p. 207).

The second issue is the cognitive benefits of bilingualism. Vygotsky's theories (1962) about the positive impact of second language learning on metalinguistic awareness and concept formation through abstraction appear to be supported by other studies on the cognitive advantages of bilingualism (Cummins, 1987; Glick, 1987; Lambert, 1990).

Hornberger's *continua of biliteracy* is the third theoretical lens. These continua of biliteracy represent an ecology of language framework for explicating biliteracy as a dynamic, multi-layered, and multi-dimensional confluence of factors impacting individual biliteracy (Hornberger, 2004, 2008; Hult, 2008; Verhoeven, 1997). Thus, the overarching question for Hornberger's model is "*What is the ecological framework for biliteracy development?*" Figure 9 presents this conceptual model of biliteracy as intersecting continua triads organized into four nested factor taxonomies. Hornberger's purpose is to "demonstrate the multiple and complex interrelationships between bilingualism and literacy and the importance of the contexts, media and content through

which biliteracy develops" (Hornberger, 2004, p.156). Thus, this model allows researchers to define biliteracy as a phenomenon and construct within specific individual and social contexts (Hornberger, 2008). For example, oral language proficiency levels in two languages (i.e., bi-oracy) might be captured as points on the three continua streams (see Figure 9) within the *development* level while reading might appear in a continuum at the *content* level.

Hornberger's model also reveals delimiting factors in other levels (e.g., media, context). In the aforementioned example, factors (e.g. linguistic properties of L1 and L2) could be identified that would predict L2 oral language proficiency levels using different instructional delivery models or approaches. Additionally, this biliteracy model can express skill transfer and growth across the continua and permit model development that more accurately depicts the second language acquisition process (Hornberger, 2008; Hornberger & Johnson, 2007).

In conclusion, *theory* more than individual research findings, "permits the generation of predictions about program outcomes under different conditions" (Cummins, 1999, p. 26). Thus, in terms of the theoretical assumptions of my study, I relied on Chomsky for an overview of the biological and linguistic origins of language development. I relied on Vygotsky for an overview of the sociocultural origins of language development and learning. And, I relied on Hornberger for an overview of the ecological origins of the biliteracy process. These cognitive, sociocultural, and ecological theories of language development facilitate an understanding of the biological

basis for linguistic behaviors and the sociocultural interactions shaping those behaviors (John-Steiner & Souberman, 1978).

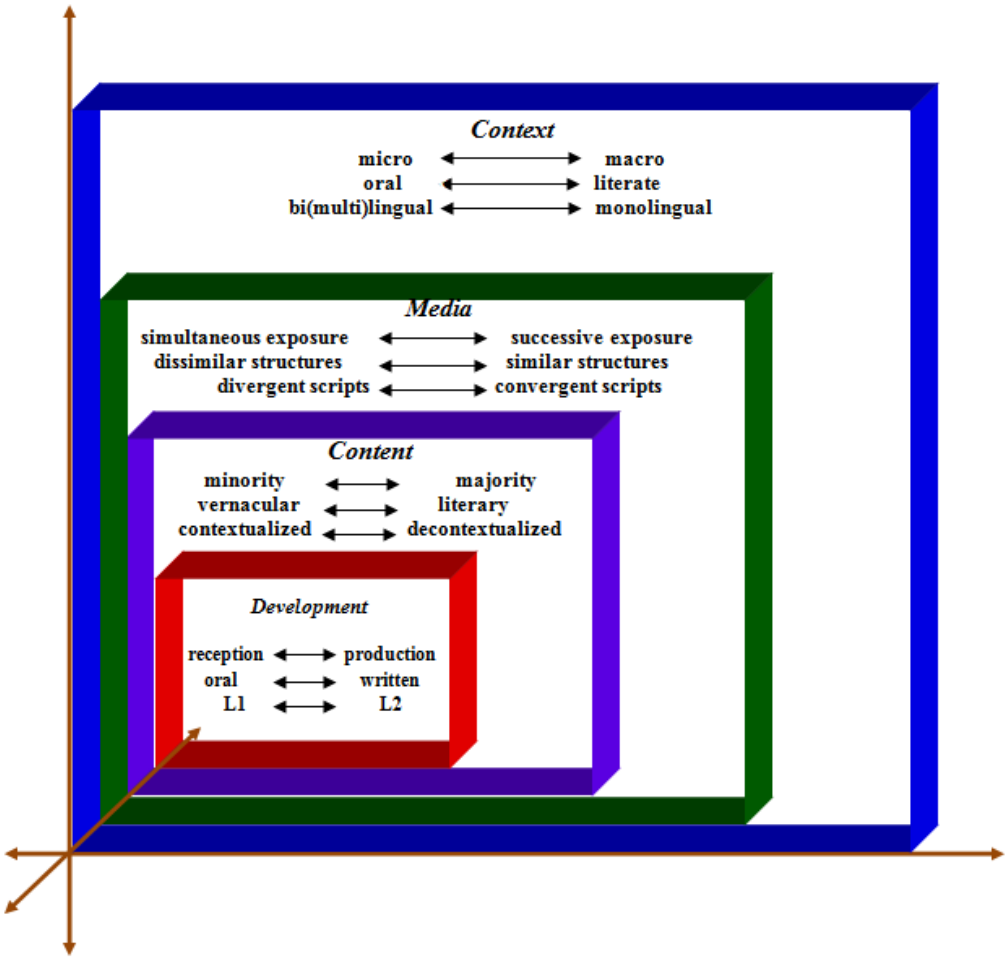

Figure 9. Continua of Biliteracy. Adapted from Hornberger, N.H. (2008).

VITA

Name: Sandra T. Acosta

Address: 107F Harrington Tower MS 4225
College Station, TX 77843

Email Address: sacosta@tamu.edu

Education: B.A., Spanish The University of Texas, Austin
M.A., Spanish University of Wisconsin, Madison
M.Ed., Public School Administration St. Thomas University, Houston
Ph.D., Educational Psychology Texas A&M University, College Station

Professional Experience: Assistant Professor, Bilingual Education Program, Educational Psychology Department, College of Education and Human Development, Texas A&M University, College Station, 2009-present

Lecturer, Bilingual Education Program, Texas A&M University, College Station, 2005-2009

Undergraduate Coordinator, Bilingual Education Program, Texas A&M University, College Station, 2005-2008

Bilingual/ESL Coordinator, La Porte ISD, 2003-2005

Assistant Principal – Jackson Middle School, Houston ISD, 2000-2003

Assistant Principal – Clemente Martinez Elementary, Houston ISD, 1998-2000

Bilingual Educational Diagnostician, Houston ISD, 1995-1998

Bilingual Education Teacher, Houston ISD, 1987-1995

Professional Affiliations: American Educational Research Association
American Psychological Association
Houston Area (Texas, National) Association(s) for Bilingual Education
Kappa Delta Pi
Southwest Educational Research Association