Examining Teacher Qualifications, Characteristics, and Practices Related to English Learners in the State of Alabama: An Exploratory Study on Teacher Quality

by

Heidi T. Goertzen

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 1, 2015

Keywords: teacher quality, characteristics, qualifications, practices,
English learners, academic language

Approved by

Lisa Kensler, Chair, Associate Professor of Educational Foundations, Leadership and Technology
Joni Lakin, Assistant Professor of Educational Foundations, Leadership and Technology
Lynne Patrick, Associate Clinical Professor of Educational Foundations, Leadership and Technology
Jamie Harrison, Assistant Professor of Curriculum & Teaching
Frances Kochan, Professor Emerita

Abstract

The purpose of this study was to explore how general classroom teachers report their qualifications, characteristics, and practices related to teaching English learners (ELs) academic language. It was also the purpose of this study to explore whether these reported qualifications and characteristics predicted teacher reported instructional practices. This research design used a survey questionnaire to measure how general classroom teachers reported (a) knowledge and skills using practices that support the development of academic language of EL students, and (b) their teaching qualifications and characteristics.

Descriptive statistics were used to assess how general classroom teachers reported on qualifications, characteristics, and practices to teach EL students academic language. Multiple linear and hierarchical multiple regression analyses were used to assess certification, hours of training in EL professional development, preference teaching the EL population by size, and experience teaching EL students as predictors of teacher reported knowledge and skills teaching academic language.

Findings from this study revealed most survey participants have had limited training and reported low levels of knowledge and skill sets related to EL students, and that certification in English as a Second Language (ESL/ESOL), over 40 hours of EL professional development, preference teaching large populations of EL students, and experience teaching at least 5 EL students in the prior school year significantly predicted and positively related to how teachers tended to report knowledge and skill level in teaching EL students academic language.

Acknowledgments

I would like to express heartfelt thanks to my committee chair, Dr. Lisa Kensler, for always encouraging and having faith in me. Through her rigorous expectations and kind, thoughtful words, I persisted and believed in myself throughout this incredible journey. I am also grateful and indebted to Dr. Joni Lakin whose statistical expertise stretched my thinking and growth as a researcher. I never knew I had it in me to navigate data in this respect. I was extremely fortunate to have Dr. Jamie Harrison's insights on English learners—a core reason for writing my dissertation. I also thank Dr. Lynne Patrick not only for her expertise in instructional leadership, but also for her human touch and caring personality. Finally, I thank Dr. Fran Kochan whose regard for craftsmanship and scholarship—and whose leadership—gave me the understanding and confidence to embark on this dissertation journey.

I also thank my dearest colleagues and friends, Shay Baugh and Amanda Inabinett, my cohort partners. Without your friendship, support, and care, writing my dissertation would have been an even greater challenge. I walked down this path from beginning to end with strength, love, and encouragement that you gave me, and I am sincerely grateful to have shared this road together.

Finally, I thank my father for instilling in me the value of hard work and a strong sense of adventure, and for my loving mother whose encouragement and faith have kept me going, seeking new adventure. And I express my deepest gratitude to my husband Leslie and my daughter Ginger Renee, to whom I owe the most love and appreciation for sacrificing so much of

our time. I thank Les for his love, support, and advice along the way, and I thank Gigi for being the most selfless, understanding, and gracious young girl. I look forward to making up for those long nights and weekends with more memorable times together wherever the world takes us. And I would be remiss if I didn't thank my four-legged friends Guero, Carmela, and Cinnamon, for their companionship and lasting joy.

Table of Contents

List of Tables

List of Figures

CHAPTER I. INTRODUCTION

Introduction

This research study examined the constructs of teacher quality as they relate to the instruction of English learners (ELs), a fast-growing population of students enrolled in US schools and who are limited in English proficiency and require placement in language assistance programs. Teachers of ELs need to use specialized instructional practices to develop their students' academic language and content as a means of promoting achievement (Calderón, Slavin, & Sánchez, 2011; Echevarria, Vogt, & Short, 2012). However, many teachers lack formal training and preparation to meet the educational needs of this population (Coady, Harper, & de Jong, 2011; Lucas, Villegas, & Freedson-Gonzalez, 2008; Reeves, 2006), while an array of teacher standards tend to gloss over the instructional needs of this special population, further contributing to the knowledge gap among classroom teachers (Jones, Buzick, & Turkan, 2013; Turkan, Croft, Bicknell, & Barnes, 2012).

In response to federal education reform on teacher quality and evaluation, many states, including Alabama, are currently revamping their teacher evaluation systems affording policymakers and practitioners the opportunity to integrate design elements that specifically address English learners. Given there is scant research on teacher quality related to ELs (Loeb, Soland, & Fox, forthcoming), the purpose of this study was to (a) identify how teachers report instructional practices they use related to the development of academic language of ELs and (b) examine the relationship among their qualifications, characteristics, and instructional practices,

constructs of a framework for teacher quality (Goe, 2007). Considering most teacher evaluation models are designed without regard to these variables, the extent to which teachers are aware and practice them is unknown. Thus, if teacher quality and evaluation systems do not establish explicit expectations for how teachers should respond to unique needs of ELs, to what extent do teachers currently address them? The state of knowledge is such that this initial exploratory study is needed.

Chapter 1 opens with background on teacher quality from a national standpoint in light of federal policy and reform efforts on teacher evaluation. Then it provides the statement of the problem, the purpose of this study, and its significance. Next, the theoretical framework for teacher quality is illustrated, followed by the three research questions. Finally, Chapter 1 concludes with the limitations of the study, the assumptions, definition of terms, and the study's overview.

Background

Teacher quality maintains its position on the forefront of education reform as educators felt its first major overhaul under the highly ambitious *No Child Left Behind Act (NCLB) of 2001*. More recent federal initiatives have expanded reform efforts, including the Race to the Top grant, the 1003(g) School Improvement Grant (SIG), and the Elementary and Secondary Education Act (ESEA) Flexibility Waiver application that permits State Education Agencies (SEAs) to waive certain *NCLB* requirements, including meeting Adequate Yearly Progress (AYP) targets and other federally funded provisions. During the economic recession in 2009, these reforms were aimed at stimulating the economy and satisfying the unanswered reauthorization of *NCLB*.

Race to the Top incentivized SEAs with $4.35 billion to reform key components of their education systems. These reform principles focus on four areas: developing more rigorous standards and assessments that prepare students for college and career; building robust data systems that indicate student growth and improve practice; turning around lowest performing schools; and recruiting, retaining, and supporting effective educators (US Department of Education, 2009). Regarding the latter initiative, Race to the Top invited SEAs to develop and implement more rigorous and fair educator evaluation systems that include the use of student growth measures for every child as a significant factor among other measures of performance. An eligibility requirement for Race to the Top included no state-level legal barriers to linking student achievement data or growth to teachers and principals as part of educator evaluation reform (US Department of Education, 2009).

In a similar vein, the SIG 1003(g) (US Department of Education, 2010) and the ESEA Flexibility Waiver (US Department of Education, 2012) incentivized SEAs with funding opportunities for turning around low-performing schools and waiving unpopular accountability targets. As such, SEAs would agree to reform their educator evaluation systems to reflect a multitude of measures including data on student growth as measured by academic assessments that culminate in a summative evaluation score. The overarching federal initiative is the ESEA Flexibility Waiver, as 43 of the 45 states that have submitted requests have been approved. Each of these federal reforms includes educator evaluation as a key principle. That said, the majority of states have either modified, or are currently revamping these models.

Specific requirements under the ESEA Flexibility waiver include evaluation models that (a) will be used to continuously improve instruction; (b) measure educator performance using three levels; (c) use multiple measures in determining performance levels including student

growth for all students—English Learners (ELs) and Students With Disabilities (SWDs)—and other measures that may include standards-based observations, student and parent perception surveys and other meaningful and valid metrics; (d) evaluate educators on a regular basis; (e) provide feedback indicating needs and strengths; and (f) use evaluation outcomes in personnel decisions (US Department of Education, 2012). With respect to the latter requirement, educator evaluation has now become high stakes for teachers and leaders.

To address the achievement gap, *No Child Left Behind* approached teacher quality in terms of teacher qualifications and equity. For example, it required teachers to be state certified in the core subject area for which they were assigned to teach, and it encouraged districts and schools to place these highly qualified and more experienced teachers in schools with low-income students. *NCLB* did not address, however, teacher quality in terms of English as a Second Language (ESL) teachers since they are not considered teachers of core-subject areas. While some empirical studies support mandates regarding highly qualified teachers under *NCLB* (e.g., Heck, 2007), other findings are unclear as to whether teacher credentials such as type of degree or certification have an impact on student achievement (Hanushek & Rivkin, 2010).

Researchers, policymakers, and practitioners agree, however, that traditional models of teacher evaluation need reform, as research indicates that teacher effectiveness is the highest predictor of student achievement (Stronge, et al. 2007). Two notable studies on teacher quality and effectiveness support this need for change. In the "Widget Effect" a historical survey of evaluation ratings indicated that in districts using binary rating systems, 99% of teachers received a rating of satisfactory (as opposed to unsatisfactory) (Weisberg, et al., 2009). This practice suggests that some models of teacher evaluation are perceived as perfunctory

compliance tools versus continuous improvement tools that facilitate recognition of teacher excellence and needed professional growth.

The second study, known as the Measures of Effective Teaching (MET) project sponsored by the Bill and Melinda Gates Foundation, in 2009 with a team of researchers and participating school districts from across the nation sought out how to identify effective teaching and ways to measure it. This study examined the use of multiple measures, including student perception surveys, value-added models, and classroom observations of teacher performance. This influential large-scale study has impacted decision-making among policymakers and practitioners, as federal policy now requires the use of multiple measures (U.S. Department of Education, 2012). When multiple measures are taken as a composite, they lend more reliability and validity to teacher evaluation outcomes in addition to providing a meaningful platform for professional growth (Kane & Staiger, 2013).

As a result of these and previous groundbreaking studies, the landscape of teacher quality is evolving toward more sophisticated, yet reliable and meaningful, systems of evaluation that use multiple sources of data including prediction models linking student growth and achievement to teacher performance (Kane & Staiger, 2013; Munoz, Prather & Stronge, 2011; Strong, Gargani, & Hacifazlioğlu, 2011; Stronge, Ward, & Grant, 2011; Stronge, Ward, Tucker & Hindman, 2007). Although controversy abounds with respect to student growth models (Baker et al., 2010), the shift to using multiple measures of teacher effectiveness presents a more accurate picture of teacher performance than status quo. Moreover, combining and weighing these measures meaningfully increases reliability and validity (Kane, McCaffrey, Miller & Staiger, 2013) on summative teacher evaluation outcomes.

The EL Subgroup and Teacher Quality

While much research sheds light on more accurate ways of measuring teacher performance, there is currently no mention of what these measures mean for teachers who serve subgroups such as English learners. Over the past two decades education reform has focused on closing the achievement gap; thus, including subgroups in the design of these new models of teacher evaluation aligns with this ongoing effort. The cornerstone of *NCLB* was to disaggregate student subgroup data as a mechanism for accountability and continuous improvement to help close the achievement gap (U.S. Department of Education, 2001).

The opportunity to design evaluation systems that address the achievement gap presents itself to states and districts as they augment current systems. Studies show that English learners benefit from unique instructional practices and particular teacher characteristics (Loeb, et al., forthcoming), yet teacher evaluation instruments typically do not measure these special indicators nor consider these other variables (Holdheide, Goe, Croft & Reschly, 2010; Jones, Buzick & Turkan, 2013). Given that teacher quality is the single highest predictor of student achievement, designing evaluation systems that integrate these variables shows promise in closing the achievement gap between ELs and non-ELs.

Statement of the Problem

Studies that examine meaningful, fair, and reliable approaches to teacher evaluation and teacher quality that are inclusive of ELs are extremely limited. As problems of reliability and validity surround both traditional and current models of evaluation for general classroom teachers (Baker et al., 2010; Kane & Staiger, 2012; Meyer, Cash, & Mashburn, 2011; Strong, Gargani, & Hacifazlioğlu, 2011), these models present unique challenges for the same teachers serving ELs. These problems are associated with standards-based evaluation and observation

instruments (Baker, Gersten, Haager & Dingle, 2006; Jones, Buzick & Turkan, 2013), student growth models (Jones, Buzick & Turkan, 2013; Lakin & Young, 2013), and demographic variables related specifically to ELs and other special populations (Buzick & Laitusis, 2010; Johnson & Semmelroth, 2014; Jones, Buzick & Turkan, 2013; Jones & Brownell, 2014; Turkan, Croft, Bicknell & Barnes, 2012).

When an evaluator observes teachers serving ELs or calculates their individual effects on student growth, careful consideration of factors unique to this subgroup has the potential to increase fairness, validity, and reliability of the assessment (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick & Turkan, 2013). Most teacher evaluation models lack these variables in their design, some of which include responding to the student's English proficiency level, home language, cultural background, age, literacy level, transiency, time in a language development program, testing accommodations, performance on achievement tests, influence from co-teachers, specially designed instruction for ELs, and other critical factors (Jones et al., 2013; Calderón, Slavin & Sánchez, 2011). The extent to which teachers are aware of these constructs and the degree to which they respond to them, both instructionally and in terms of evaluation, is unknown.

Purpose of the Study

As states and school districts in Alabama redesign teacher evaluation systems to reflect more fair and accurate measures, it is relevant to explore the constructs of teacher quality as they relate specifically to the EL subgroup. The primary purpose of this study was to examine these constructs through the reporting of general classroom teachers and to explore the relationships among these constructs. Specifically, this study's aim was to examine the qualifications and characteristics of general classroom teachers in Alabama who serve in schools with 20 or more

ELs and how they report their skill level in terms of the practices they use to develop academic language of ELs.

Significance of the Study

This study sought to explore evaluation systems that can promote teacher growth and accountability while providing the same high standards for ELs. Researchers argue that observation protocols matter with regard to student achievement (Baker et al., 2006; Gallagher, 2004; Gersten et al., 2005; Heneman, Milanowski, Kimball, & Odden, 2006; Milanowski, 2004), thus, evaluation designs that integrate the unique academic and linguistic needs of the EL population inform researchers, policymakers, and practitioners of effective teaching practices that show promise for closing the achievement gap and increasing graduation rates among ELs (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick, & Turkan, 2013; Master, Loeb, Whitney, & Wyckoff, 2012). The more scientifically-based evidence that exists to inform the educational community on how teachers affect the achievement and progress in academic language development of ELs, the closer it will be to developing evaluation models that are equitable, fair, valid, and reliable tools for measuring teacher effectiveness for all children (Samson & Collins, 2012). Moreover, it would serve to stimulate further interest and research into this topic.

Theoretical Framework

Multiple approaches to defining and measuring teacher quality exist. Scholars have conceptualized teacher quality to provide a framework for conducting empirical research; however, they have yet to arrive at consensus (Blanton, Sindelar & Correa, 2006). The framework for teacher quality is evolving, with scholars defining the construct in terms of teacher experience and characteristics to behavior and action linked to student outcomes and

successes. Current trends focus on linking teacher performance to student outcomes, as the prevalence of student achievement tests facilitates this measure.

To examine the constructs specifically related to the instruction of ELs, this study used Goe's (2007) framework for teacher quality. Goe conceptualizes teacher quality through a framework of three interrelated strands: (1) inputs, (2) teacher processes, and (3) outcomes. The inputs are the qualifications and characteristics teachers bring with them to the classroom. These inputs feed into the next strand, processes, which are the practices teachers use within the context of the school and classroom. Finally, these two strands impact student outcomes, namely, student achievement scores, which, under this particular framework, are indicators of teacher quality. Goe (2007) classifies these student gain scores as teacher effectiveness, a concept that is "empirically defined using value-added measures, [wherein] teachers are ranked by how much students gained compared to how much they were predicted to gain in achievement" (p. 9). It is noteworthy to mention, however, that scholarly opinions differ with regard to the definition of teacher effectiveness (Blanton et al., 2006; Kennedy, 1999).

Constructs related to characteristics of teachers of ELs include teacher (a) attitude toward linguistically and culturally diverse students and families (e.g., Reeves, 2006), such as preference to teach at schools where there is a high concentration of ELs (Loeb et al., forthcoming), and (b) attributes, such as the shared language and culture of their students. Meanwhile, teacher experience, training, and certification represent constructs of teacher qualifications. Next, the construct of teacher practices involve use of instructional skills, such as (a) teaching academic language and second language literacy (Calderón, Slavin & Sánchez, 2011), (b) identifying the language demands of the content and explicitly teaching these functions and structures, (c) integrating student background and culture in curriculum and instruction, and (d) applying

knowledge of second language acquisition concepts in order to differentiate instruction (De Jong, & Harper, 2005; Lucas, Villegas, & Freedson-Gonzalez, 2008). Systems of teacher evaluation that integrate these constructs can potentially render the design of student growth models and standards-based observation rubrics more valid and reliable while serving as measures of overall effectiveness.

Research Questions

Compelling questions emerge from the literature on teacher quality concerning interpretations of the instructional quality provided to ELs. "If observation protocols are to support valid interpretations about teachers' instructional quality for ELs…researchers must attend to whether observers themselves can reliably differentiate between teachers who do and do not make use of effective instructional practices for [ELs]" (Jones, et al., 2013, p. 238). Moreover, research is needed to gauge perceptions of practitioners on EL teacher quality in order to examine its current standing. As a first step in exploring this problem, analyzing the perceptions of general classroom teachers of ELs provides data that add to the research base on teacher quality and English learners. The following research questions address these concerns.

1. What number of factors in the *Teacher Quality* survey is identified through exploratory factor analysis?

2. To what extent do teachers report that they are prepared for teaching ELs academic language?

3. To what extent do teachers report they have adequate skills to teach ELs?

4. To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number of ELs taught in the classroom predict teacher reported skill level for teaching academic language?

Limitations of the Study

The length of the survey may have contributed to attrition. In future studies the survey instrument could be shortened since only one factor was extracted through the Exploratory Factor Analysis, addressing non-completers. In terms of representativeness, determining how sample participants reflect the population sample can only be measured by the demographic data they self-reported, such as certification area, teaching assignment, race, gender, and age, as the survey was anonymous.

Assumptions of the Study

There is evidence to suggest that participants were honest in their survey responses. There was good variation in responses with teachers reporting a broad range of having somewhat low to very high skills in teaching academic language. Also, there was a sufficient percentage of teachers who reported not preferring to teach a large group of English learners, which appears to indicate honest responses versus reporting what would be the correct responses.

Definition of Terms

Academic language – A construct used to define the type of linguistic knowledge and skills students need to be successful in a school setting. Researchers and theorists consider this type of register beneficial for college or career opportunities (Scarcella, 2003).

English Learners (ELs) – Students whose proficiency in English is limited and who receive English language development services.

English as a Second Language (ESL) – A term used to describe teachers of English learners, as well as language development programs for ELs.

Teacher quality – The term used in research on teacher effectiveness and evaluation.

Overview

Chapter 1 described the background for the study and introduced the problem statement, the purpose of the study, and its significance. A discussion of the theoretical framework was provided to focus the study. Three research questions were posed, followed by limitations and assumptions. Chapter 2 provides a comprehensive review of the literature related to the topic of study, while Chapter 3 describes the research design that corresponds to the three research questions. A description of the sampling methods, participants, data collection instruments, and analysis techniques is presented. Chapter 4 describes the data analysis and results, followed by the final chapter, which presents a discussion of the findings that emerge from the analysis, along with implications for policy and practice and recommendations for future studies.

CHAPTER II. REVIEW OF THE LITERATURE

Teachers are key contributors to student achievement (Jones & Brownell, 2014; Stronge et al., 2007). For example, students who spend a year with an ineffective teacher require at least three years to regain standard levels of achievement (Mendro, 1998), whereas those who have been taught by an effective teacher for just one year tend to outperform their peers in subsequent years. Statistically, "A one-standard-deviation improvement in teacher effectiveness (going from the average teacher to one at the 84th percentile) would move the average student from the 50th to the 56th percentile in the year with the better teacher" (Hanushek & Rivkin, 2010, p. 134).

Meanwhile, school administrators tend to place disadvantaged students, usually those who fall under multiple demographic subgroups, with teachers who are less effective at increasing student growth (Betts, Zau, & Rice, 2003; Borman & Kimball, 2005; Sanders et al., 1997), which contributes to the achievement gap. Researchers who conducted groundbreaking studies on teacher evaluation systems conclude that reform is necessary for judging teacher performance in more valid, reliable, and meaningful ways, both in terms of improving student learning and equity, and for teacher recognition and growth (Kane & Staiger, 2012; Kane et al., 2013; Weisberg et al., 2009).

Measures of teacher evaluation that differentiate teacher performance outcomes are beneficial in many ways. For example, school administrators who are able to distinguish between high and low performing teachers can make teaching assignments and placement more equitable with regard to low-income and at-risk students. Second, outcomes from these

evaluation measures indicate a pathway for teacher growth and can provide insight into what quality teaching looks like, as researchers and administrators alike are able to learn about the differences in processes that occur within those classrooms (Muñoz, Prather & Stronge, 2011; Muñoz, Scoskie & French, 2013; Strong, 2007). Surprisingly, there is a paucity of research on teacher quality and the achievement of specific subgroups, namely English learners (Loeb, Soland & Fox, forthcoming; Master, Loeb, Whitney & Wyckoff, 2012), given the role *NCLB* played with regard to disaggregation of test data and closing the achievement gap. The examination of teacher quality as it relates to the EL subgroup is a missing piece (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick, & Turkan, 2013; Turkan, Croft, Bicknell, & Barnes, 2012).

Nationally, English learners (ELs) are the fastest growing student population, with the most rapid growth seen in grades 7 through 12 during 2007–2008 (Calderón, Slavin, & Sánchez, 2011). Between 1994 and 2010 the growth rate of ELs nationally was 64%, up from 3.2 to 5.2 million EL students (NCELA, 2011). The National Center for Education Statistics reported 9.1% of students in U.S. schools participated in programs for English learners (4.4 million) in 2011–2012. While the numbers fluctuate, there is clearly an upward trend. Many states not accustomed to serving students who are limited English proficient have experienced a sharp influx over the past decade although educators in these states continue to lack the knowledge and skills to provide the type of instruction these students need to be successful academically (Lucas, Villegas, Freedson-Gonzalez, 2008).

English learners present challenges to teachers in terms of their unique pedagogical and linguistic instructional needs (Turkan, Croft, Bicknell, & Barnes, 2012), and data indicate that ELs are not responding to current approaches to instructional services in terms of student

outcomes. For example, the achievement gap was over 20% on the NAEP math and reading tests for 4th and 8th grade students in 2013. Table 1 illustrates these gaps in detail. While much of the measures and standards of teacher quality pertain to general classroom practices, attention to the instructional practices associated with the achievement of ELs and teacher quality is extremely lacking (Jones, Buzick, & Turkan, 2013; Loeb, Soland, & Fox, forthcoming; Lucas et al., 2008; Master, Loeb, Whitney, & Wyckoff, 2013; Samson & Collins, 2012).

Table 1

Summary of 2013 NAEP Achievement Gaps Between ELs and Non-ELs who Scored Proficient on Math and Reading Tests

Student Demographic	4th Grade Math %	8th Grade Math %	4th Grade Reading %	8th Grade Reading %
ELs	13	4	6	3
Non-ELs	36	28	29	33
Achievement Gap	23	24	23	30

(Adapted from NAEP, the Nation's Report Card, 2013)

The remainder of Chapter 2 is divided into two parts and will discuss the literature on teacher quality and evaluation with regard to English learners. The first part pertains to the framework for teacher quality (Goe, 2007), beginning with an overview of its three constructs. This section will then include a comprehensive review of the research on teacher inputs, processes, and outcomes, both in terms of general education and the EL subgroup. The second part of Chapter 2 bridges teacher quality to teacher evaluation reform models. Toward that end,

a brief overview of current approaches to evaluating teachers will be provided, followed by a discussion of the gaps in evaluation designs. Finally, three research questions will be presented to address the knowledge gap among researchers, policymakers, and practitioners regarding teacher quality, evaluation, and English learners.

Teacher Quality

Goe (2007) conceptualized teacher quality using the following three strands: (1) inputs (teacher qualifications and teacher characteristics), (2) processes (teacher practices), and (3) outcomes (teacher effectiveness). This conceptual framework allows for a systematic examination of the domains related to teacher quality; however, it is important to note that while the framework facilitates the discussion of the literature, the elements of each strand are complex and do not necessarily fall neatly into one category. There is overlap. For example, the outcomes strand is prominently featured in a majority of the more recent studies on teacher quality (Baker, Gersten, Haager, & Dingle, 2006; Muñoz, Prather, & Stronge, 2011; Strong, Gargani, & Hacifazlioğlu, 2011; Stronge, Ward, & Grant, 2011) as numerous research designs use the measure of teacher effectiveness—student growth—as part of its methods in objectively identifying differential teacher effectiveness, which leads to the exploration of inputs and processes teachers use.

It is important to clarify that while this study applies all three constructs to frame the literature on teacher quality, it examines the relationship only between the first two, inputs and processes. The third strand, outcomes (teacher effectiveness) will not figure into the design of the research study. Figure1 illustrates Goe's (2007) conceptual framework for teacher quality and delineates the scope of the research design.

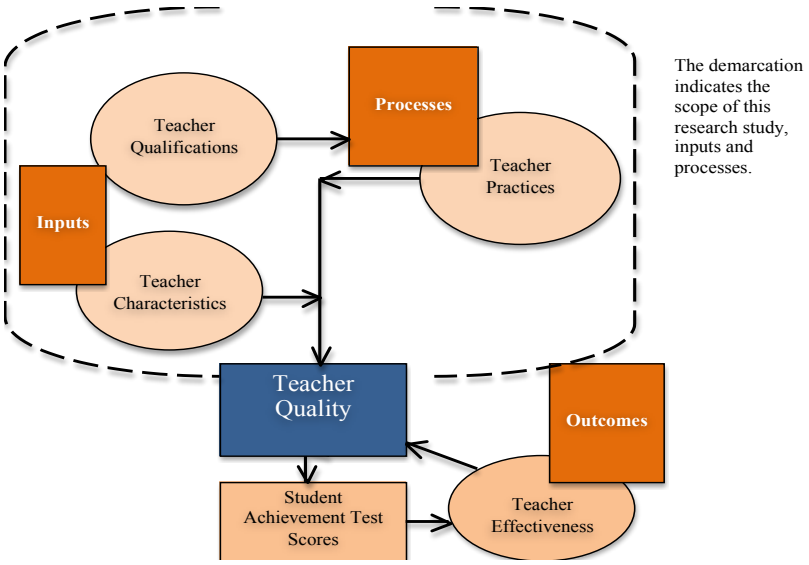

Figure 1. Framework for Teacher Quality (Goe, 2007).

Inputs

Teacher qualifications. The category of teacher qualifications, which falls under inputs, focuses on teacher credentials such as type of degree, years of experience, type of professional development, continuing education units, license exam scores, and other paper-based inputs. Administrators generally employ these variables in the hiring and placement of teachers, especially as a result of *NCLB's* highly qualified teacher legislation. Researchers indicate that math teacher credentials are more significant at the secondary level, but have moderate to low significance at the elementary levels (Cavalluzzo, 2004), and less is known about other subject area credentials (Goe, 2008). Hanushek and Rivkin (2010) assert that there is little evidence to support a strong correlation between teachers who have earned a master's degree and increased student achievement.

Through a slightly different approach, Heck's (2007) study examined teacher quality in terms of certification and teaching assignment as mandated by *NCLB*, which requires teachers to

be highly qualified as defined by the state education agency. Heck (2007) examined teacher qualification at the school level (as opposed to individual teachers) and its relationship to student achievement; this study found evidence in favor of *NCLB*'s mandate that teachers must be certificated (highly qualified) in the subject areas for which they are assigned to teach. However, these findings are inconsistent with those of Betts et al. (2003) who examined a variety of teacher qualifications within the San Diego Unified School District and found their correlation to student achievement varied substantially.

Master, Loeb, Whitney and Wyckoff (2012) examined the relationship between several teacher qualification variables and achievement of ELs in mathematics by comparing them to the same teachers' non-EL outcomes. In comparing these two groups, researchers found a weaker relationship between teacher competency test scores and EL math achievement and no significant differences regarding teacher experience and math achievement between both groups. However, their findings indicate that prior experience teaching at least six ELs in the classroom is highly predictive of math EL achievement with implications that more practice working with ELs in the classroom is beneficial. In addition, these scholars found that training and experience with instructional practices unique to ELs "predict differential effectiveness with ELL students, particularly among novice teachers" (Master et al., 2012, p. 25). In terms of teacher experience in general, researchers have found that teacher performance improves over time, albeit only during the first few years of teaching; thereafter, improvement plateaus (Clotfelter, Ladd & Vigdor, 2007; Nye, Konstantopoulos, & Hedges, 2004).

Teacher characteristics. Teacher characteristics involve teacher attitudes and attributes, including race and gender. Studies conducted within this domain do not provide sufficient evidence of impacts on student achievement, thus more research is needed in this area (Goe,

2007). However, a recent study explores the relationship between predictor variables of bilingual teachers and teacher disposition (among several other variables) with EL achievement in math (Loeb et al., forthcoming). The variable for teachers who speak the students' first language predicted greater growth in math among ELs than non-ELs. However, these findings were inconsistent with a previous study (Master et al., 2012) that found no significant relationship between teachers who share the student's native language and student achievement in math. This finding merits further examination of effects on reading achievement of ELs since math was the only dependent measure used in this study—a step taken in Loeb et al. (forthcoming).

A key finding in the Loeb et al. study is that the variable for motivation, i.e., teachers with a preference for working with the EL subgroup, was predictive of teacher effectiveness in math. Peercy's (2011) qualitative study on secondary-level ESL teacher instructional practices also indicates some evidence that sharing the same background such as race and language improves the teacher's level of cultural responsiveness (Peercy, 2011). To illustrate, while one of the teachers in this study did not share the same race or language as her ELs,

> [She] questioned her effectiveness because she did not share the same background as her students, but her awareness of this difference also seemed to make her work harder to understand her students' perspectives. She recognized herself as a cultural being and realized that her culture and her students' cultures affected the ways in which they interacted and learned, which is critical to culturally responsive teaching. (Peercy, 2011, p. 347)

Toward that end, investigating culturally responsive teachers as a predictor of EL achievement could fill a gap in the research on teacher characteristics as inputs.

Considering the vast array of research designs and methods, studies on teacher characteristics produce inconsistent findings (Goe, 2007). Regardless, some are compelling and warrant more research, such as Dee's (2004) findings, which indicate that when placed with a teacher of the same race, positive student achievement among Black students was cumulative. Another example is a study by Ehrenberg, Goldhaber, and Brewer (1995), who found evidence suggesting that teacher gender may affect the way they evaluate their students.

In sum, findings from research that explores both general education and EL teacher characteristics are inconsistent, and thus evidence is unclear regarding effects on achievement for all students. What is known thus far about EL teacher characteristics is that certain variables may influence EL student achievement (Loeb et al., forthcoming; Master, et al., 2012; Peercy, 2011). While, language, ethnicity, culture, and motivation are factors that have been explored in terms of teacher quality and ELs, more research is needed as these findings come from only two known studies.

Processes

Teacher practices – general classroom teachers. Numerous studies explore teacher practices using value-added measures to identify quality instruction (Baker, Gersten, Haager, & Dingle, 2006) and high and low performing teachers (Muñoz, Prather, & Stronge, 2011; Strong, Gargani, & Hacifazlioğlu, 2011; Stronge, Ward, & Grant, 2011). The application of student growth modeling to explore teacher practice exemplifies how the three strands of the teacher quality framework (Goe, 2007) interact. For instance, student growth models are equivalent to the outcomes strand in the framework, and researchers are looking at those outcomes as a way to identify quality teaching. A prime example is the Muñoz et al. (2013) study, which examines the "black box" of effective teaching.

Muñoz and his research team (2013) used hierarchical linear modeling (HLM) to develop a classroom academic index to identify high and low effective teachers. Once identified, teachers from each group were randomly selected to participate in a survey using constructs from Stronge's (2007) theoretical framework for teacher quality. Researchers collected teacher perceptions of the characteristics of effective teachers through a survey instrument. Data from the survey suggest that more effective teachers tended to value providing a physically and emotionally safe environment, whereas less effective teachers placed a higher value on limiting interruptions and focusing on teaching and learning. In general, however, both groups tended to agree on most effective teaching indicators posed in the survey. A limitation to this study was the exclusion of diverse learners. As noted by the authors, "future research may consider inclusion of all students (e.g., special needs, English Language Learners) in elementary schools; this would offer an opportunity to better understand the "black box" of effective classrooms for diverse students" (Muñoz et al., 2013, p. 226).

Researchers conducting similar studies (Stronge et al., 2007; Stronge, Ward, & Grant, 2011) inquire on the differences in instructional practices teachers with high and low student growth levels use. In the earlier study, Stronge et al. (2007) employed a mixed methods approach, first using HLMs to identify effective and ineffective teachers with student learning gains as the dependent variable. Once these teachers were identified, they conducted case studies to compare these two groups' instructional practices. Findings show three notable differences between teachers who "effected greater than expected learning gains for students…(1) differentiation and complexity of instructional strategies, (2) questioning practices, and (3) level of disruptive student behavior" (p. 180).

The first finding indicated that effective teachers in this study were more successful at employing a variety of instructional strategies that met students' individual needs. And while both groups of teachers used approximately the same number of lower level questions, the effective teachers used seven times more higher-order questions. The difference in performance of classroom management observed in this study was also dramatic, with effective teachers tending to disruptions once every three hours compared to once every twelve minutes for the ineffective teachers (Stronge et al., 2007).

In their second study, Stronge et al. (2011) also explored behaviors of effective and ineffective teachers, inquiring on the degree to which teachers have positive effects on student achievement in math and reading. Again, HLMs were used to gauge effective and ineffective teachers for the first phase of the study, and the second phase involved teacher observations to explore behaviors and a perception survey to assess teacher beliefs and efficacy. Results found that:

> Students taught by bottom-quartile teachers could expect to score, on average, at the 21st percentile on the state's reading assessment, whereas students taught by the top-quartile teachers could expect to score at approximately the 54th percentile. This difference, more than 30 percentile points, can be attributed to the quality of teaching occurring in the classrooms during one academic year. (p. 344)

Findings from this study suggest that teachers who have a strong rapport with their students and strong classroom management skills with few disruptions outperform teachers in the low percentiles as measured by HLMs.

Because these studies replicate previous work, they succeed at distilling key indicators of effective teaching beliefs and practices in the mainstream classroom—the inputs and processes:

(a) providing an emotionally safe learning environment, (b) frequent use of higher order thinking questions, (c) complex and differentiated instruction, and (d) classroom management skills. Key teaching methods and teacher attitudes are integral to the processes that lead to positive student learning gains. Granted, other important indicators exist and are part of the teaching equation, but knowing which ones tend to show more promise toward growth in student achievement is highly relevant. While these studies do not directly address teacher evaluation, they target measureable practices that comprise teacher standards and teacher quality.

A caveat to consider with regard to studies that use HLMs or similar types of regression modeling to assess teacher effectiveness centers on how these models are designed. For example, variables used to build these models have limitations. Multiple years of data are more reliable predictors of teacher effectiveness than one year of data (Loeb & Candelaria, 2012), but often their availability is lacking, which was the case in the latter study; in addition, random sampling of students is practically impossible to attain when attributing teacher performance to student growth, as strategic placement of students is a central process administrators and teachers tend to use in practice (Braun, 2005; Loeb et al., forthcoming). For research purposes, however, using HLMs has utility for identifying variance in teacher performance.

While using linear regression models to discern effective from ineffective teachers is instrumental in current studies on teacher quality, they do not explain the processes and procedures teachers use that contribute to their level of teaching effectiveness. Studies that delve further into the "black box" add to the knowledge base on what effective—and ineffective teachers do. Muñoz, et al. (2013) call for this deeper inquiry —"we need to uncover the 'black box' of actual student learning by studying the real actors. We need to incorporate teachers and student voices in the important debate around teacher effectiveness and evaluation" (Muñoz, et

al., 2013, p. 228), hence, their use of a perception survey that probes these practices and beliefs. A question that remains, however, is: what are teacher and administrator perceptions of practices and characteristics of effective teachers of ELs?

Teacher practices – EL teacher practices. Research on teacher effectiveness and the black box of classroom practice has not been conducted as fervently, if at all, on EL teacher quality as it has for general education teachers (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick, & Turkan, 2013; Turkan, Croft, Bicknell, & Barnes, 2012; Turkan, de Oliveira, Lee, & Phelps, 2014). Rather, researchers who study ELs focus on interventions and treatments that improve EL student learning in various subject areas (Turkan, Croft, Bicknell, & Barnes, 2012) and that promote progress in academic language acquisition and second language literacy (August & Shanahan, 2006; Baker, Gersten, Haager, & Dingle, 2006; Baker, et al. 2014; Calderón, Slavin, & Sánchez, 2011; Echevarria, Vogt, & Short, 2012; Gersten, Baker, Haager, & Graves, 2005; Lucas, Villegas, & Freedson-Gonzalez, 2008; Peercy, 2011; Téllez, & Waxman, 2005). Some scholars have examined effectiveness of EL teacher preparation programs (Coady, Harper, & de Jong, 2011; Lucas, Villegas, & Freedson-Gonzalez, 2008), while others have analyzed indicators of language services and program effectiveness (Calderón, Slavin, & Sánchez, 2011). Tying these indicators of effectiveness to EL teacher quality as it pertains to evaluation is lacking (Samson & Collins, 2012; Turkan, Croft, Bicknell, & Barnes, 2012).

Effective teaching practices for ELs may not have a niche carved out in current teacher quality studies related to teacher evaluation; however, extant research and theory do provide constructs of EL teacher quality so the road to evaluation can be paved. Researchers and practitioners assert that principals and evaluators of teachers of ELs should be familiar with effective EL teacher practices and skills including (a) second language acquisition (SLA), (b)

student background and culture, (c) providing access to the content area, (d) using visuals and gestures, (e) differentiating instruction based on student's English proficiency level, (f) communication with parents, and (g) use of the student's first language (L1) to develop the second language (L2) (National Comprehensive Center for Teacher Quality, 2012).

Theoretical and empirically based constructs in applied linguistics serve as the foundation for teaching and learning language, and principles of second language acquisition underpin these constructs. Thus, teacher knowledge of SLA is linked to instructional practices for ELs, which in turn, impacts lesson planning—a case of inputs (qualifications) impacting processes (practice) in terms of the teacher quality framework. Some concepts that fall under the SLA umbrella include (a) the stages of language acquisition (b) the use of the student's native language (L1) to acquire the second language (L2), (c) comprehensible input (such as using visuals and gestures), and (d) factors affecting SLA, such as student's English proficiency level, language distance (i.e., how closely linked the L1is to the L2, e.g. Italian and Spanish), rate of language acquisition, L1 literacy, and affective filter (Echevarria, Vogt, Short, 2012).

Academic language and literacy for ELs. Researchers have examined a multitude of variables and interpreted their implications for effective instruction of ELs, which has led to the constructs that undergird language teaching and that are relevant to teacher quality. These studies tend to focus on two major areas: (a) practices impacting the acquisition of academic language and content and (b) effective literacy practices for learning to read in a second language. Surveying the literature that examines these variables provides a knowledge base from which to link them to EL teacher quality in terms of evaluation purposes. Before delving into research on these two domains, laying the groundwork on academic language is an essential precursor to the discussion. Academic language is a multi-faceted dimension of language

acquisition (Gottlieb & Ernst-slavit 2014; Scarcella, 2003), and debate exists as to a solid definition (Kieffer & Lesaux, 2012; Lesaux, 2010). Scholars have attempted to conceptualize it in various ways.

In his seminal work, Cummins (1979; 2000) introduced a theory coined Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP). This linguist posits that individuals learn BICS with—to an extent—minimal effort and efficiency, as it falls under the social realm of communication. In simple terms, BICS is the social language of the playground, school cafeteria, and other less cognitively demanding environments where meaning is highly contextualized and linguistic cues are easily identifiable. Under this linguistic framework language learners acquire BICS earlier than academic language, and aspects of it are taught in English language development programs, especially for recent non-English speaking enrollees—consider survival guides for newcomers that explain where bathrooms are located, provide directions to various school locales, and other helpful information.

Acquiring CALP, on the other hand, takes much longer, seven or more years depending on an individual's literacy background and related contextual factors (Collier, 1987; Collier & Thomas, 1989). CALP is the type of language needed to navigate through cognitively demanding, abstract concepts typically found in academic settings. All students will have to reckon with acquiring proficiency in cognitive academic language upon entering school, and the level of linguistic sophistication only increases at a commensurate rate with grade levels (Uccelli et al., 2014); for low-income, at-risk, and language minority students (students who come from homes where a language other than English is spoken), acquiring CALP is an even greater

challenge as these subgroups generally enter school with low language development levels (Lesaux, 2010).

Scarcella (2003) warned, however, that to embrace the dichotomous view of BICS and CALP prevents language educators from seeing the broader scope of literacies that are not confined to "the highly structured English associated with school-based tasks…" (p. 5). Scarcella presented a framework that underscores the complexities of academic language and how intricate its development is:

> Academic English includes multiple, dynamic, inter-related competencies. It provides a compilation of a broad range of discrete linguistic items so that teachers and researchers are provided with sufficient information concerning what the language features of academic English are. It also provides information concerning the psychological, social, and cultural factors associated with academic English. It suggests that without detailed information, teachers have difficulty teaching their students and assessing their academic English proficiency. (p. 7)

Current literature on the construct of academic language focuses on building teachers' conceptual understanding and recommends processes teachers of ELs can use to explicitly promote academic language in core-subject areas, including math, science, social studies, and English/Language Arts. As such, teachers have the potential to develop their students' proficiency in this type of register (Hwang et al., 2014; Turkan et al., 2014). Hwang et al. (2014) argued that because academic language uses "complex language structures, has a higher proportion of low frequency vocabulary, includes nominalization of verbs and abstract nouns, and makes less use of personal pronouns…" (p. 3), providing students with a robust academic vocabulary skill intervention would facilitate acquisition of this academic register.

Further, Turkan, de Oliveira, Lee, and Phelps (2014) made a direct connection between EL teacher quality and effective teaching practice in the context of academic language and System Functional Linguistics (SFL). Resonant of BICS and CALP (Cummins, 2000), they describe academic language as the language of schooling that is distinct from informal language, and they support the notion that effective teachers are able to identify the language demands of the content area and develop their student's awareness of the difference (Anstrom as cited in Turkan et al., 2014, p. 6–7). The second strand, System Functional Linguistics reflects work by Halliday and Hasan (1989) and Schleppegrell, (2004), scholars guided by the tenets of sociolinguistics and metacognition. Operating under SFL, teachers aim to develop their students' awareness of the appropriateness of the type of language needed for certain contexts. "Knowing the discourse of a discipline, therefore, is a form of socialization into the ways the members of the discipline talk, write, and participate in the knowledge construction by making the appropriate linguistic choices to convey the meaning" (Turkan et al., 2014, p. 8).

Using the lenses of academic language and SFL, Turkan et al. (2014) proposed the Disciplinary Linguistic Knowledge (DLK) framework, the essential knowledge and skill that teachers of ELs must possess in order to engage their students in the language of their discipline. Under this framework, teachers build their students' understanding and accurate use of oral and written language specific to their discipline. In particular, DLK expects teachers to be able to (a) identify linguistic features of the disciplinary discourse and (b) model how to communicate meaning in the discipline and engage students in using the language in all domains (listening, reading, writing, and speaking) (Turkan et al., 2014, p. 9).

Despite differences in how academic language is defined, viewed, or approached, scholars agree that English learners and language minority students must be able to access and

produce language found in academic settings in order to be successful, and determining the most promising instructional practices is essential for academic success (Carlo et al., 2004; Cummins, 2008; Echevarria, Vogt, & Short, 2012; Kieffer & Lesaux, 2012; Lesaux, Kieffer, Kelley, & Harris, 2014; Scarcella, 2003; Snow & Uccelli, 2009; Nagy & Townsend, 2012). Moreover, these instructional practices characterize EL teacher quality; they are the criteria that define good teacher practice. For quality instruction for ELs to occur and be measured, teachers must be able to translate these concepts of academic language into concrete practice, and researchers are exploring how best to teach it in light of student outcomes and its implications for policy and practice. The following section addresses an empirical body of research on academic language and literacy.

Academic language and literacy practices. Most recently, a panel of researchers reviewed and synthesized robust research studies that focused on teaching academic language and literacy to ELs (Baker et al., 2014). As a result, the research team provided four instructional recommendations for kindergarten through eighth grade teachers in the following areas: (a) intensive instruction of academic vocabulary, (b) development of oral and written English skills integrated with content teaching, (c) development of written language skills, and (d) small group instruction (Baker et al., 2014). These recommendations are based on a number of vetted empirical designs that qualified for inclusion in the research synthesis and are based on specific constructs of language teaching and learning from which researchers have found positive effects on student outcomes. As Master et al. (2012) assert in their study on EL teacher quality and effectiveness, "more research is needed to better understand which specific skills are most relevant for supporting academic achievement among English language learners" (p. 27).

A list of the various studies on academic language and literacy and the respective recommendations based on rigorous studies (Baker et al., 2014) is provided in Table 2. The sections that follow are discussions of the research conducted in each of the four areas of effective academic language and literacy instruction.

Table 2

Recommendations for Development of Academic Language and Literacy in Grades K–8

Findings/ Recommendations	Research to Support Findings
Intensive Instruction in Academic Vocabulary	Carlo et al., 2004; Kieffer & Lesaux, 2012; Lesaux, Kieffer, Faller, & Kelley, 2010; Lesaux, Kieffer, Kelley, & Harris, 2014; Silverman, 2007; Silverman & Hines, 2009
Development of Written Language Skills with Content Teaching	Kim et al., 2011; Gomez, Parker, Lara-Alecio, & Gomez, 1996
Oral and Written English and Content	Ballenger, 1997; Brown, Ryoo, & Rodriquez, 2010; Echevarria, Vogt, & Short, 2012; Gersten, 1996; Ryoo, 2009
Small Group Instruction for Students who Struggle	Bradley et al., 2011; Burns, 2011; Denton, Wexler, Vaughn, & Bryan, 2008; Nelson, Vadasy, & Sanders, 2011; Ransford-Kaldon et al., 2010; Solari & Gerber, 2008; Vaughn et al., 2006

Adapted from the Research Synthesis Compiled by Baker et al. (2014).

Academic vocabulary. English learners benefit from vocabulary instruction designed for Non-ELs, and evidence suggests they may benefit even more (August & Shanahan, 2006; Carlo et al., 2004; Silverman & Hines, 2009). Much of the research on effective vocabulary instruction

has been conducted with English speakers in mind (e.g., Beck, McKeown, 2007; Graves, 2000; Nagy, 1997), and second language researchers investigating academic vocabulary advance these studies to determine their effects on ELs (Carlo et al., 2004; Kieffer & Lesaux, 2012; Lesaux, Kieffer, Faller, & Kelley, 2010; Lesaux, Kieffer, Kelley, & Harris, 2014; Silverman, 2007; Silverman & Hines, 2009). These scholars use findings from past studies (Baumann, Kame'enui, & Ash, 2003; National Reading Panel, 2000) as treatment variables in designs that focus on English learners and language minority students' academic vocabulary development. These robust designs on ELs are randomized and compare groups of ELs with Non-ELs, determine which variables predict student outcomes in reading and vocabulary, and yield results on effect sizes (Carlo et al., 2004; Kieffer & Lesaux, 2012; Lesaux, Kieffer, Faller, & Kelley, 2010; Lesaux, Kieffer, Kelley, & Harris, 2014).

Findings from research on ELs and academic vocabulary (Carlo et al., 2004) indicate that teaching academic words, developing awareness of polysemy (words with different but related meaning), teaching the use of strategies to infer meaning from unknown words, and instruction in morphological awareness and cross-linguistic word meaning had positive effects on student performance for both groups. A critical finding in the Carlo et al. (2004) study suggests that effective vocabulary instruction centers on in-depth word study versus teaching a high volume of vocabulary as a means to maximize language input. As a result, learners are provided with tools that enable them to independently decipher meaning based on their ability to understand word parts (morphology). Carlo et al.'s (2004) study served as a precursor to current and future studies on effective instruction of academic vocabulary for ELs.

For example, Lesaux et al. (2010) expand on the work of Carlo et al. (2004) using quasi-experimental, mixed methods to determine if there are differences in outcomes between ELs and

Non-ELs although this study employs different instructional methods and contexts. The intervention, Academic Language Instruction for All Students (ALIAS), focuses on building word knowledge skills using a framework that moves chronologically, with an eight-day cycle, first introducing high-utility words from a carefully selected, high-interest text to expecting students to know how to use the word precisely in writing. The findings were consistent with Carlo's (et al., 2004) study, with similar effect sizes on target word mastery, knowledge of meaning of words in context, and morphological awareness. These findings also suggest that ELs should be taught smaller numbers of vocabulary words using high interest texts and in-depth instruction that includes multiple meanings, and morphological analysis, a predictor of reading comprehension skills. The treatment used in this study had positive effects on developing vocabulary and comprehension skills of middle school ELs.

To extend the Lesaux (2010) study, Kieffer and Lesaux (2012) focused on a narrower angle with unused data from the prior study. They investigated the effects of the ALIAS intervention on EL and Non-EL performance on relational and syntactical aspects of morphological skills, i.e., performance on real-word decomposition (e.g., knowing how and when to remove a suffix) and non-word derivation (e.g., understanding meanings of non-words when a suffix is added, as in "or" that connotes a person, *trentor*). The researchers found that ELs outperformed non-ELs on non-word derivation, with possible implications that the intervention is uniquely advantageous for ELs. With a sharper focus on morphology, Keiffer and Lesaux (2012) concluded that morphological awareness (a) should occur in a systematic scope and sequence, (b) should be explicitly taught as a metacognitive strategy, and (c) it should be taught using rich instructional context that uses high-utility academic words with multiple opportunities for learning and using words.

The notion of high-utility words plays a significant role in the intervention posed in research-based academic vocabulary instruction (Baumann & Graves, 2010; Hwang et al., 2014; Nagy & Townsend, 2012). These academic vocabulary words are found throughout the curriculum, in different subject areas; examples include *conflict, increase, propose, and analyze*. They do not include content-based words such as *perpendicular, organism, latitude, or alliteration*.

Other researchers explore vocabulary interventions that place a heavy emphasis on concepts of second language acquisition (Gersten, Baker, Graves, 2005; Silverman, 2007; Silverman, 2009), including comprehensible input (Echevarria, Vogt, & Short, 2012; Krashen & Terrell, 1983). For example, Silverman (2007) investigated the differences in overall growth of vocabulary instruction using growth modeling between kindergarten level ELs and non-ELs. This researcher used similar treatments from prior "mainstream" studies and augmented them with instruction designed specifically for ELs. For example, students had the opportunity to act out word meanings, and they were provided with illustrations of vocabulary words. Findings were consistent with previous studies on academic vocabulary instruction (Carlo et al., 2004) and indicate that ELs learn vocabulary as fast as or faster than non-ELs, suggesting a potential means to close the achievement gap between those groups in word knowledge.

In a different study on academic vocabulary, Silverman (2009) found that using a multimedia enhanced read-aloud vocabulary intervention does not provide an added benefit for non-ELs when compared to the use of a read-aloud without the multimedia. However, the researcher found that this approach did have positive effects on EL outcomes, suggesting use of multimedia enhanced read-aloud vocabulary instruction is possible for inclusive settings. Researchers (Gersten et al., 2005; Silverman 2007, 2009) have found that adjustments to

instruction that include visual aids facilitate learning academic vocabulary for ELs and aligns with principles of SLA such as comprehensible input.

Oral and written English and content. The empirical studies related to academic vocabulary discussed above describe the study of words per se, with some explanation of how to locate these words, the quantity of words to use, and type of text recommended for instruction. The latter two vocabulary research designs, however, explored practices related to SLA that involve kinesthetic activities and visual media, which align with instructional concepts recommended when teaching cognitively demanding, abstract academic language (Cummins, 2008). Sound instructional design for language learners maintains cognitive demand of the content combined with highly contextual strategies and activities that provide students with access to the material (Cummins, 2008). Toward that end, researchers have found that instruction that aptly addresses this notion of cognitive demand and context develops written and language skills in the content areas for ELs (August, Branum-Martin, Cardenas, Hagan, & Francis, 2009; Ballenger, 1997; Brown, Ryoo, & Rodriguez, 2010; Cummins, 2008; Echevarria et al., 2012; Fuchs, Fuchs, Mathes, & Simmons, 1997; Gersten, 1996; Ryoo, 2009).

August et al. (2009) effectively address this approach in their QuEST study, which uses as its treatment, a sixth grade level science-based curriculum that incorporates instructional strategies designed for ELs that make the content comprehensible. In this cluster randomized controlled experiment, the research team made adjustments to the science curriculum by integrating scaffolding techniques, such as visual aids, graphic organizers, experiments, modeling, and demonstrations so the content of the science lessons was more accessible for ELs. This intervention included increased opportunities for interaction and discussion about the science topics as teachers clarified and elaborated on student responses. Findings from this study

indicate growth among ELs and non-ELs alike. Notably, this research team operates under the principle that instructional interventions for one group of students must not be deleterious for the other, given that in the real world, classrooms are typically heterogeneous. In this case, both groups benefitted.

Findings from the Ryoo (2009) study, which also focused on science education and academic language, indicate positive effects for student outcomes in both content and language when students are introduced to a scientific concept using everyday English before employing a more scientific register during instruction (Ryoo, 2009). This process aligns with Disciplinary Linguistic Knowledge (DLK) framework that Turkan et al. (2014) advance in regard to EL teacher quality. In the Ryoo (2009) study, however, ELs benefitted more from the treatment variables than did their English-speaking counterparts, suggesting there are some instructional strategies that support academic language acquisition for those who need it while bearing neutral effects for those who do not. Again, these outcomes align with the philosophy that instructional interventions for one group of students must not be deleterious for the other (August et al., 2009).

Brown, Ryoo, and Rodriguez (2010) extend the study of academic language in science education by integrating the construct of disaggregate instruction and discursive identity in their design. Disaggregate instruction "is the idea that science teaching and learning can be separated into conceptual and discursive components" (p. 1465). To illustrate, similar to Ryoo's (2009), study, treatment entailed the use of everyday English to discuss concepts as a precursor to using intensive academic English to discuss these same topics. Yet in this study, Brown et al. (2010) delve further and propose that learners are hindered by their deficiency in the scientific register

to engage in discourse on scientific phenomenon due to their particular affiliation with a different cultural group or subculture. They refer to this membership as discursive identity.

Brown et al. (2010) found positive effects on student outcomes in the treatment group, and recommend scaffolding instruction by first teaching concepts in ordinary language as a means of supporting learning, and talking about science in a manner that students can understand, then transitioning to a more formal, scientific register. They refer to this approach as disaggregate instruction. The findings from the study also underscore socio-cultural and sociolinguistic (Halliday & Hasan, 1989; Schleppegrell, 2004) aspects of SLA and implications of affect (Krashen & Terrell, 1983). Toward that end, Turkan et al. (2014) argue that teachers must make the distinction between everyday language and the language of the content being taught explicit to their students. There are two ideas at play in this case: (a) the teacher is tapping into the metacognitive aspect of teaching and learning, and (b) using everyday English is considered a type of scaffold for language learners.

A recurring theme throughout these studies on academic content and language development involves comprehensible input and scaffolding instruction, important elements of sheltered instruction (Echevarría, Vogt, & Short, 2012). Sheltered instruction is an approach that makes content comprehensible to ELs in the mainstream classroom while simultaneously developing their academic language. One approach to sheltered instruction developed in the 1990s is known as the Sheltered Instruction Observation Protocol (SIOP) Model (Echevarría, Vogt, & Short, 2012). Originally, SIOP was a research tool, designed to measure the degree to which teachers were able to implement sheltered instruction strategies that comprise the model, but it later evolved into a comprehensive tool for instruction of ELs and is currently a widely used professional development tool (Short, Fidelman, & Louguit, 2012). While the SIOP model

was not an explicit element of the four recommendations for teaching academic content and literacy (Baker et al., 2014), its foundation, sheltered instruction, is an integral part of developing academic content and literacy; moreover, its application extends beyond grades K–8. A description of this model and the research that has been conducted on it is warranted.

The model consists of eight indicators of instruction, with a total of 30 items, that when implemented to fidelity, effect improvement on student outcomes in language and literacy (Echevarría, Short, & Powers, 2006; Short, Echevarría & Richards-Tutor, 2011). The eight indicators of instruction found in the SIOP Model include (a) lesson preparation, (b) building background, (c) comprehensible input, (d) strategies, (e) interaction, (f) practice/application, (g) lesson delivery, and (h) Review/assessment (Echevarría, Vogt, & Short, 2012). Thirty features comprise the eight indicators and include items such as vocabulary instructional strategies, the inclusion of content and language objectives, appropriate age and educational background level content concepts, scaffolding and higher order thinking skills, teaching cognitive and meta-cognitive learning strategies, and others. Since its inception, researchers (Echevarría, Richards-Tutor, Chinn, & Ratleff, 2011; Short, Echevarría, & Richards-Tutor, 2011; Short, Fidelman, & Louguit, 2012) have studied the effects of the SIOP Model on student outcomes in achievement and its effectiveness as a professional development model.

For example, in their quasi-experimental design Short, Fidelman, and Louguit (2012) examine the effects of sheltered instruction on the academic literacy performance of secondary school ELs as well as the degree of teacher ability to implement the model after one and two years of training in the SIOP Model. Outcomes from this study indicate positive effects on students' English language development as measured on instruments assessing writing, oral language, and total English scores, and that the model is a predictor of these three areas in

student achievement. Short, Fidelman, and Louguit (2012) note that the SIOP Model "improved the quality of teaching in the content classes" (p. 353), a key point that links to the processes strand of the teacher quality framework.

Echevarría et al. (2011) conducted a similar study exploring the influence of teacher fidelity to the SIOP Model on student growth in content area literacy in seventh grade science classrooms. In this randomized control study of eight middle schools in a large urban district, researchers found that teachers who had higher SIOP scores implemented the SIOP Model to a higher degree and had higher student gains than low scoring teachers. Most notably, this study found that teachers with high implementation levels of the SIOP Model used a wider variety of interactive strategies that provided students with more opportunities to practice using academic language. They also provided more comprehensible input, and modeled expectations for students, whereas low-scoring teachers provided a more teacher-centered style of instruction and used sheltered instruction strategies less frequently.

Researchers found that the level of teacher fidelity to the SIOP model impacted student outcomes. A variety of explanations are possible for the differential implementation of the SIOP Model, including the design of the study, and other plausible variables, but teacher quality cannot be ruled out (Echevarría et al., 2011). "A confound that must be considered is the possibility that the highest implementers were simply the best teachers" (Echevarría et al., 2011, p. 433). Toward that end, Echevarría et al. (2011) assert that as districts and schools implement policy on teacher evaluation designs, implementing an observation rubric with a continuum of performance is an important consideration as teacher outcomes in this particular study were based on the degree and frequency of implementation versus the presence or absence of the instructional indicators.

Findings from the Echevarría et al. (2012) study parallel those conducted by Stronge et al. (2007) on teacher effectiveness insofar as complexity and variety of instructional strategies serve as indicators of quality instruction and quality teaching. Further, overall findings from these SIOP studies (Echevarría, Richards-Tutor, Chinn, & Ratleff, 2011; Short, Echevarría, & Richards-Tutor, 2011; Short, Fidelman, & Louguit, 2012) compare with studies on the integration of oral and written language with content teaching, which strongly suggest that when students are provided with frequent opportunities to practice and use academic language, learning outcomes improve (August et al., 2009; Brown et al., 2010; Ryoo, 2009).

Provide structured opportunities to develop written language skills. Although a dearth of robust scientific studies on teaching practices in academic writing for ELs and language minority students have been conducted, what is known about this process is that a student's English proficiency level has much to bear on writing outcomes (Kim et al., 2011; Short & Fitzgerald, 2007). Until English learners have acquired enough basic language skills, they will be less successful at engaging in more complex academic language practices such as expository writing (Short & Fitzgerald, 2007).

Kim et al. (2011) found that a students' level of English language proficiency impacts their ability to successfully engage in particular cognitive learning strategies for writing analytical text related to literature practices in secondary level English language arts classes. In this study researchers found a relationship between a student's English proficiency level and the type of cognitive writing strategy being taught, consistent with earlier research on language learning strategies (e.g., Chamot, 2004). For example, an EL who is at the beginning stages of language proficiency does not have sufficient command of the language to engage in certain types of cognitive learning strategies for writing essays.

Students at these early levels of English proficiency may benefit more from visual aids along with a high degree of context-embedded supports such as graphic organizers to scaffold writing, whereas cognitive strategy sentence starters would suffice for intermediate level ELs (Kim, et al., 2011). Because of the heterogeneity of ELs in terms of their language proficiency level, cognitive strategies used to scaffold their learning, as well as develop learner autonomy, require differentiation (Chamot, 2004; O'Mally & Chamot, 1995; Salahshour, Sharifi, & Salahshour, 2013).

Approaches to instructional writing practices for language learners, including the study of transitional words (such as *however, although, in terms of*), the use of routines to scaffold learning, differentiating instruction based on a student's English proficiency level, and features of process writing that are linked with evidence-based vocabulary instruction skills (e.g., Lesaux et al., 2014), underscore the intricacy of acquiring academic language. When teaching ELs, all pieces of the language puzzle must be addressed in a coherent scope and sequence as language is, by default, integrative (Lesaux et al., 2014); thus, each aspect of effective instructional practice for ELs constitutes the fabric of academic language development (Scarcella, 2003; Téllez & Waxman, 2005). These effective practices for teaching academic language and literacy are models of English language development instruction. However, when some students do not respond to this core instruction, more intensive approaches are needed (Mathes, Pollard-Durodola, Cárdenas-Hagan, Linan-Thompson, & Vaughn, 2007; Vaughn & Linan-Thompson, 2003; Vaughn, et al., 2006). These methods are discussed in the next section.

Small group instruction and reading in a second language. In addition to academic vocabulary, the integration of language and content, and the development of writing skills, another significant element of effective classroom practices for ELs includes second language

reading instruction (August & Shanahan, 2006; Baker, et al., 2014; Gersten, et al., 2005; Lesaux & Siegel, 2003). Approaches to reading and literacy instruction for ELs vary to an extent, and may depend on grade span levels, i.e., elementary or secondary (Denton, Wexler, Vaughn, & Bryan, 2008). Thus researchers tend to focus their questions on literacy acquisition for students who are either elementary or secondary level ELs, with the latter centering on mainly on middle school, given that high school ELs face an altogether different set of literacy challenges (Baker et al., 2014; Short & Fitzsimmons, 2007). The discussion that follows describes various studies on literacy for ELs at both elementary and middle school levels, but it focuses in particular on effective literacy instruction for students who consistently struggle with acquiring academic language and literacy.

Much of what is known about teaching reading to English speakers in the early grades is true for teaching ELs (Baker et al., 2006; Gersten, Baker, Haager, & Graves, 2005; Gersten et al., 2005; Linan-Thompson, & Vaughn, 2007; Mathes, Pollard-Durodola, Cárdenas-Hagan, Lesaux & Siegel, 2003). A wealth of research on teaching reading in the early grades shows that schools implementing critical early reading principles, including phonemic awareness, graphophonemic knowledge, word recognition, fluency, vocabulary, and comprehension, are effective in terms of reading achievement in the primary grades (Griffin, Burns, & Snow, 1998; Lesaux & Siegel, 2003; National Reading Panel, 2000). However, teachers who implement early literacy practices need to adjust instruction for the EL population by employing strategies that make learning new content more comprehensible (Gersten, et al., 2005; Short, Fidelman, & Louguit, 2012; Silverman, 2007; Silverman, 2009; Solari & Gerber, 2008).

Yet a significant problem educators face is the number of students, both EL and non-EL, who struggle with reading despite the use of these early reading interventions. Aligning

instruction to these key reading principles is not sufficient for struggling readers, as more time and intensive instruction is required (Mathes, Pollard-Durodola, Cárdenas-Hagan, Linan-Thompson, & Vaughn, 2007; Vaughn & Linan-Thompson, 2003; Vaughn, et al., 2006). The situation for native Spanish speaking ELs who struggle with reading is compounded when both their English and Spanish language proficiency levels are low (Carlo, et al., 2004).

To achieve greater instructional intensity, studies show that placing students in small groups to focus on explicit literacy skills instruction improves performance in many of the literacy domains (Burns, 2011; Denton, Wexler, Vaughn, & Bryan, 2008; Nelson et al., 2011; Solari & Gerber, 2008; Vaughn et al., 2006). A variety of studies on ELs and language minority students have been conducted, each focusing on slightly different interventions to improve reading outcomes for struggling readers, yet each with a focus on explicit, small group instruction. The framework for many of these studies mirrors Response to Intervention, or *RtI* (Mathes, et al., 2007; Nelson et al., 2011), a tiered approach that uses small group instruction as a supplement to core classroom instruction when students do not positively respond to instruction of the school's core reading program (Denton & Mathes, 2003; Linan-Thompson, Vaughn, Prater, & Cirino, 2006; Vaughn, Linan-Thompson, & Hickman, 2003).

Small group instruction for young ELs. Vaughn et al. (2006) conducted a randomized control study on Spanish speaking first graders in Texas who were at-risk for reading difficulties, where participating schools offered Spanish bilingual instruction. The language of the intervention used in the study matched the language of the core reading program for each participating school. So for some treatment and control groups the intervention was in Spanish while for others it was in English.

Treatment included small group instruction with attention to explicit instruction on phonemic awareness, phonemic decoding skills, fluency in word recognition and text processing, construction of meaning, vocabulary, spelling, and writing. Because students in the study were ELs with low language proficiency skills in both Spanish and English, researchers prioritized oral skills and vocabulary development. Teachers did not use direct instruction during instruction that focused on these linguistic skills; rather dialogue was included as a means to engage students in conversation with the content and vocabulary words used in text.

Vaughn et al. (2006) concluded "that treatment group students performed significantly higher than comparison students on critical outcomes measured in Spanish, including phonemic awareness, word attack, word reading, reading comprehension, fluency, and overall language ability in Spanish" (p. 68). Moreover, students in the treatment group made significant gains in most skill elements and were better prepared for second grade, making a smooth transition into an English only reading program. A notable explanation Vaughn et al. (2006) offer for such growth is the possible impact of the integration of oral language and vocabulary instruction with the other literacy skill areas. Another factor that may have contributed to growth is the design's sampling method, whereby participating schools were selected based on their high performing record defined by the researchers as "the requirement that their EL learners had 80% pass rates or higher on the state reading assessment in third grade" (p. 69).

In a similar study on early literacy acquisition, Solari and Gerber (2008) conducted a randomized alternate treatment control group study on early comprehension instruction for Spanish speaking kindergarten level ELs at risk for reading difficulties. These researchers explored listening comprehension (LC) and phonemic awareness (PA) skills as a model of instruction for teaching early literacy (Solari & Gerber, 2008). In this study, researchers defined

Listening Comprehension as "listening to connected, meaningful passages with instructional dialogue" (Solari & Gerber, 2008, p. 165) with text skill subsets that included summarization, main idea, recall, predicting, and making inferences. Phonemic Awareness included skill subsets for decoding (rime/onset detection and segmentation/blending). The researchers based their hypothesis on prior studies suggesting that early LC instruction is a strong predictor of later reading comprehension (Aarnouste, van den Bos, & Brand-Gruwel, 1998; Duke, Pressley, & Hilden, 2004; Proctor, Carlo, August, & Snow, 2005). Further, Solari and Gerber (2008) questioned a long held theory suggesting that comprehension skills should be developed after decoding and fluency skills are mastered.

Solari and Gerber (2008) investigated oral comprehension strategies designed for young ELs and their ability to apply these skills to new text. Remaining consistent with prior research on small group instruction (Gersten & Geva, 2003; Vaughn et al., 2006), student groups were limited to four or five students who were either at risk or not at risk. The research team examined how varying amounts of time could be more efficiently applied during instruction in order to cover different combinations of PA and LC interventions and whether there would be performance gains in PA, given the additional instructional piece on LC.

Dosage of literacy skills for each of the three groups differed. For example, the first group which focused on PA Concentration received instruction on PA 70% of the time with 10% devoted to alphabetic knowledge, and 20% to LC and vocabulary, whereas the LC Concentration group received LC and vocabulary instruction 70% of the time, with alphabetic knowledge at 10% and PA at 20%. The treatment control group received PA only, with 20% alphabetic knowledge and 80% word-level skill. Dosages of each treatment variable are listed in Table 3.

Table 3

Treatment Variable Dosages for Literacy Skills

Interventions	Phonemic Awareness (PA) %	Alphabetic Knowledge %	Listening Comprehension (LC) & Vocabulary %
PA Concentration	70	10	20
LC Concentration	2	10	70
Control Group	100	0	0

Findings indicate improvement in PA gains when the LC Concentration intervention was the main focus during literacy instruction (70% of the time). Solari and Gerber (2008) conclude that ELs who are both at risk and not at risk can receive LC intervention before they learn decoding and fluency skills without cost. "If we are to avoid merely layering new instructional targets on existing ones, new instructional models are needed that permit simultaneous teaching of multiple reading skills in ways that are both efficient as well as effective" (Solari & Gerber, 2008, p. 166). Aware of unintended consequences that can occur when applying an intervention to a particular group of students, they offer a solution that considers cost-benefit in terms of outcomes for all students (e.g., August et al., 2009).

Supplemental, small group instruction is a cornerstone of Response to Intervention, and investigating this framework of instruction with regard to the EL subgroup is the focus of many studies in literacy acquisition. For example, Nelson, Vadasy, and Sanders (2011) focused on tier two supplemental root word vocabulary and decoding intervention with Spanish speaking ELs in kindergarten to determine proximal and distal effects (directly and indirectly linked to intervention, respectively) of the supplemental, small group instructional intervention. This

design was a randomized trial that used early vocabulary connections as its treatment and a modified form of interactive book reading as the control. Treatment was provided in a pull-out setting, 20 minutes per day, 5 days a week, from October through April.

Nelson et al. (2007) systematically vetted and developed a high frequency word list to use in the supplemental literacy intervention for this study. They based their approach on prior work that suggests ELs are not as familiar with high frequency words as their English-speaking counterparts (Beck, McKeown, & Omanson, 1987; Nation & Waring, 1997), and that early decoding literacy instruction in mainstream classrooms assumes a typical students' lexicon consists of the high frequency words used to teach decoding (Nelson et al., 2007). The researchers hypothesized that the supplemental root word intervention would allow students to make connections between their word study and the core reading program taught in their classrooms.

Results from this study indicate students in the treatment group significantly outperformed the control group on measures of root word vocabulary and word reading skills. High effect sizes of interest were a result of the randomized design and achievement-based measures (Nelson et al., 2011). "Although tentative, the relatively large effect size found for word reading suggests that there may be an advantage in linking supplemental vocabulary instruction with practice in phonics skills that students are being taught in the core reading program" (p. 201).

These discussions on early literacy research regarding ELs maintain that explicit, small group instruction that focuses on literacy skills in phonemic awareness integrated with principles of language acquisition, i.e., vocabulary, oral language development (Lesaux & Siegel, 2003; Vaughn et al., 2006), and listening comprehension (Solari & Gerber, 2008) show promise for

students learning to read in a second language, especially for those who demonstrate reading challenges. In practice it is possible for educators to incorporate additional interventions without cost to other evidence based reading approaches when ELs are included in a school's population (Solari & Gerber, 2008).

Small-group instruction for older ELs. Researchers investigating older ELs at the middle school level with severe reading difficulties describe similar interventions that dovetail with those designed for younger language learners (Denton, Wexler, Vaughn, & Bryan, 2008). To illustrate, researchers sought out to determine the effectiveness of a multicomponent reading intervention for middle school students with severe reading challenges (Denton et al., 2008). The intervention was a phonics based remedial reading intervention with an integrated literacy design that included sheltered instruction, vocabulary instruction, fluency, and comprehension strategies. However, the study was compromised due to excessive student absences and disciplinary infractions that resulted in in-school suspension, placement in an alternative school, or sitting in the principal's office. Results from the study indicate improvement in only one area, sight word efficiency, for both groups at a statistically significant level, with a small effect size (*Eta* squared .104).

"In this study, students' minimal response to intervention may have been related to a complex pattern of habits, self-efficacy beliefs, language development, and learning difficulties" (Denton et al., 2008, p. 85). Gottardo (2002) and Miller et al. (2006) contend that oral language proficiency in the primary grades for both English and Spanish speakers supports reading in both languages. Thus, Denton et al. (2008) proposed that a more focused approach on oral language development, in addition to increased intensity and duration of the intervention, may have shown more promising student outcomes. Further, the research team suggested that students "may have

had a stronger response to instruction that placed less emphasis on word identification skills and more on vocabulary and reading comprehension" (Denton et al., 2008, p. 86). English learners at the middle school level have a dual challenge of developing a second language in addition to learning the content taught in English, which is compounded by their progression through higher grade levels where language demands increase commensurately with denser text (Uccelli et al., 2014).

In sum, approaches to literacy development that integrate small group, explicit instruction for struggling readers learning English as a second language show promise when the design includes (a) oral language development such as listening comprehension that can occur before students are exposed to text-based comprehension and interaction and dialogue among students and teachers, (b) word skill development that not only allows for word meaning, but also for the instruction of morphology, and (c) special scaffolding that includes principles of sheltered instruction that makes abstract concepts more concrete (including skill-based concepts of phonemic awareness and word study).

Research shows that ELs perform as well or better than non-ELs when provided with effective early literacy practices that are integrated with principles of second language acquisition. Moreover, many research designs on ELs and early literacy use the Response to Intervention (RtI) model to address the needs of students who struggle with reading challenges. Schools implementing RtI to address the needs of their ELs who are consistently non-responsive to evidence-based instruction can make adjustments that reflect those described above to small group interventions without cost to their English-speaking counterparts.

Summary of effective practice in academic language and literacy. Studies on EL teacher quality are extremely limited (Jones, Buzick, & Turkan, 2013; Loeb, et al., forthcoming;

Master et al., 2012; Turkan et al., 2014). Rather, robust research designs on EL instruction that focus on teacher practice typically investigate second language instructional methodology using control or comparison groups to determine effective interventions for developing academic language and literacy of ELs and language minority students. Researchers who focus on processes of teacher quality, however, tend to examine teacher practices using standards-based rubrics on teacher behavior and classroom instruction, as well as their level of effectiveness based on student growth measures (Goe, 2007). Teacher quality studies that focus on the EL subgroup have not examined the practices outlined and discussed in the last section of this chapter. Below is a summary of these practices.

The most recent recommendations of evidence-based teacher practices for ELs include (a) intensive academic vocabulary instruction, (b) integrating oral and written language instruction with content teaching, (c) structured opportunities to develop written language skills, and (d) explicit and intensive small group instruction, especially for struggling second language readers (Baker et al., 2014). In terms of vocabulary instruction, teaching a small number of high-utility words, developing awareness of polysemy, providing in-depth vocabulary instruction that spans several days (Carlo et al., 2004; Lesaux et al., 2014), and teaching awareness of morphology (Kieffer & Lesaux, 2012) support the growth and development of EL and language minority students' academic language. This framework for vocabulary instruction includes making adjustments that allow for visual support (Gersten, Baker, Graves, 2005; Silverman, 2007, 2009) and comprehensible input (Echevarria, Vogt, & Short, 2012; Krashen & Terrell, 1983) as integral elements.

Integrating academic language with content concepts involves intensive vocabulary instruction, as the latter aides in comprehension and production of academic text and oral

language used in the classroom setting (August et al., 2009; Lesaux et al., 2014). Using sheltered instruction strategies such as scaffolding techniques, including visual aids, graphic organizers, experiments, modeling, demonstrations, and interaction with other students and with teachers provides ELs with access to the curriculum and frequent opportunities for oral and written language practice (August et al., 2009; Ballenger, 1997; Brown, Ryoo, & Rodriguez, 2010; Cummins, 2008; Echevarria et al., 2012; Gersten, 1996; Ryoo, 2009). Moreover, these same strategies are needed to develop academic writing skills in addition to focusing on cognitive and meta-cognitive learning strategies that are appropriate for a student's English proficiency level (Chamot, 2004; Kim et al, 2011; O'Mally & Chamot, 1995; Salahshour, Sharifi, & Salahshour, 2013). An instructional framework that promotes the integration of language skills with content offers an evidence-based foundation from which to build a knowledge base for effective EL teacher practice (Turkan et al., 2014).

Small group instruction for ELs and language minority students, wherein teachers explicitly teach early literacy skills of phonemic awareness, reinforced with sheltered instruction, oral language, and intensive vocabulary skills indicate positive effects on achievement outcomes for struggling second language readers (Denton et al., 2008; Nelson et al., 2011; Solari & Gerber, 2008; Vaughn et al., 2006). Most of this research focuses on children in the early grades although research findings on older learners are consistent with findings on literacy practices when the intervention prioritizes oral language and vocabulary development. Moreover, second language reading instruction designed for older ELs with "severely impaired word reading skills" (Denton et al., 2008, p. 86) may require a heavier emphasis on those skills in addition to comprehension, dialogue and academic conversation, and less emphasis on word identification skills (Denton et al., 2008).

In light of these evidence-based approaches to effective instruction for ELs, teacher performance in these areas has yet to be examined through the lens of teacher quality. "Despite advances in research on teacher evaluation…there has been virtually no attention given to whether teachers are effectively educating exceptional populations—namely students with disabilities (SWDs) and English learners (ELs)" (Jones et al., 2013, p. 234). A dearth of research on teacher quality related to ELs aligned with the teacher quality framework (Goe, 2007) examines variables that include teacher practices, characteristics and qualifications (Gallagher, 2002; Loeb et al., forthcoming; Master et al., 2012), and outcomes as they relate directly to the field of English language development.

For example, Master, Loeb, Whitney, and Wyckoff (2012) and Loeb, Soland, and Fox (forthcoming) have begun exploring teacher quality related to ELs using value-added models as a cornerstone of their research designs. Master et al. (2012) were concerned with effects of teacher inputs including training, experience, credentials, and language background, and their predictability of possible differences between EL and non-EL math achievement outcomes; in short, they were looking at differential teacher effectiveness using survey data as predictor variables (i.e., the inputs). While this study captures salient information on three aspects of EL teacher quality—qualifications, characteristics, and teacher effectiveness—its attention to classroom practice was a missing piece in terms of the black-box (Munoz et al., 2013) of instruction and the processes strand of Goe's (2007) framework.

It is important to note, however, that Master et al. (2012) found that training specifically related to EL instructional strategies "predict[s] significant differential efficacy in ELL math instruction" (p. 21). In addition, they found that math teachers who participated in the study and who were certified in ESL also demonstrated differential effectiveness with ELs. These findings

lead to the question of what quality teachers are doing in the classroom that have positive effects on EL student achievement. As Master et al. (2012) contend:

> Research that attends to the learning gains of ELL students in particular and that examines specific instructional interventions over time to directly assess improvement in teacher effectiveness with ELL students could inform instructional decisions and reduce the gap in achievement between English learners and other students. (p. 28)

Outcomes: Teacher Effectiveness

The third strand in the teacher quality framework (Goe, 2007) centers on teacher effectiveness using student growth measures. Student growth models are "used to determine teachers' contributions to students' test score gains" (Goe & Stickler, 2008, p. 19). Currently, a majority of states and school districts are implementing Value-Added Models (VAMs) and other forms of student growth models, including Student Learning Objectives, as per the ESEA Waiver requirements (U.S. Department of Education, 2012). As part of teacher evaluation reform, the waiver requires using multiple measures of teacher performance, one of which includes student growth as a significant factor.

Student growth models. Growth models can be classified into two groups—observed and predicted score models (Buzick & Laitusis, 2010). Models that employ observed scores are structured with a time-bound accountability target with benchmarks along the way to compare actual scores that reveal a student's trajectory toward the ultimate target. Predicted score models employ statistical methods such as linear regression, quantile regression, and multilevel modeling. As the label denotes, the latter models generate predicted outcomes based on past student achievement scores or proficiency levels.

Value-Added Models (VAMs) use sophisticated statistical models to estimate educator contribution toward student achievement (Braun, 2005; Buzick & Laitusis, 2010; Jones, Buzick, & Turkan, 2013; Loeb & Candelaria, 2012). A variety of these models exist, from the "relatively straightforward fixed effects model to a relatively complex and general multivariate, longitudinal mixed-model" (Rubin, Stuart, & Zanutto, 2004, p. 2). There are many advantages to using VAMs, as researchers argue they are able to objectively measure and isolate a teacher's contribution to student learning gains (Braun, 2005). VAMs are also better gauges of teacher evaluation compared to using measures such as observed student achievement scores or the percentage of students scoring proficient on standardized tests. School leaders may also use VAMs to identify teachers who may benefit from professional development based on their low value-added scores (Loeb & Candelaria, 2012).

Limitations of student growth models. However, VAMs also have significant limitations; for example, a necessary key assumption with regression modeling is randomized sampling (Baker et al., 2010; Rubin et al., 2004). This type of sampling is near impossible to accomplish in school settings since it is common practice for instructional leaders to deliberately place students with certain teachers as a strategy to best support students. Braun (2005) warned that "in the absence of randomization, causal interpretations can be misleading" (p. 3), yet VAMS make causal interpretations about statistical estimates. Loeb et al. (forthcoming) assert, however, that carefully designed VAMs can reduce bias, which mitigates non-random teacher assignment.

Regardless of pros and cons, policymakers require the implementation of student growth models across states and districts as part of teacher evaluation reform (U.S. Department of Education, 2012). Further, the ESEA Waiver requirements include using student growth models

to determine personnel decisions with regard to teacher evaluation systems. In light of these requirements, researchers outline specific challenges and caveats associated with VAMs (Loeb & Candelaria, 2012):

> (a) A teacher's value-added score in one year is partially but not fully predictive of her performance in the next; (b) Value-added is unstable because true teacher performance varies and because value-added measures are subject to error; (c) Two years of data do a meaningfully better job at predicting value than just one; (d) A teacher's value added in one subject is only partially predictive of her value added in another, and a teacher's value added for one group of students is only partially predictive of her value added for others; (e) The variation of a teacher's value added across time, subject, and student population depends in part on the model with which it is measured and the source of data that is used; (f) Year-to-year instability suggests caution when using value-added measures to make decisions for which there are no mechanisms for re-evaluation and no other sources of information. (p. 2)

Research designs that use VAMs are subject to the myriad limitations just described. For example, Stronge (2007) and Muñoz et al. (2011, 2013) employed Hierarchical Linear Modeling (HLMs) to distinguish between effective and ineffective teachers using student test scores. In the Stronge (2007) study, researchers identified teacher variation and explored classroom practices through observations of these teachers using the domains of instructional expertise, student assessment, learning environment, and personal qualities of the teachers. This design, however carefully constructed, lacked a sufficient sample size (N = 11), compromising generalizability and leaving room for error. Muñoz (2011) used test scores as the primary measure of teacher effectiveness, which was problematic, as using a sole test score as a

dependent measure oversimplifies the process of teacher evaluation. Moreover, using VAMs to identify effective from ineffective teachers does not shed light on the daily processes that occur inside these teachers' classrooms; other measures are needed to know what they do and how well they do it (Muñoz et al., 2011, 2013; Stronge, 2007).

Muñoz et al. (2013) also used HLMs to identify effective from ineffective teachers, yet only one year's worth of test data was built into the model. In reality, availability of data is a challenge many state, district, and school leaders face, especially since many states are in the midst of adopting new assessments that align with the Common Core State Standards. Using multiple years of achievement data yields a more robust model (Loeb & Candelaria, 2012).

Kane, McCaffrey, Miller, and Staiger (2013) also used VAMs in their Measures of Effective Teaching (MET) study, a robust design that used random assignment, significantly large sample sizes, and more than one year of test scores; further, they used a subsample of students and measures other than the state achievement test to determine teacher impact, and found overall positive effects. However, even with a more robust design that addressed the limitations described in Muñoz, Prather, and Stronge (2011), Muñoz, Scoskie and French (2013), Stronge, Ward, Tucker and Hindeman (2007), and Kane et al. (2013) describe limitations of prediction error in their study:

> Many of the classrooms taught by teachers in the bottom decile in the measures of effectiveness saw large gains in achievement; in fact, some bottom decile teachers saw average gains larger than those for teachers with higher measures of effectiveness. (p. 3)

The caveat here is the chance for error when tying a measure of effectiveness to a single teacher, especially if it involves high stakes decision-making.

Limitations of growth models to ELs. All research designs are subject to limitations, just as the ones described above that use VAMs, yet it is imperative to consider those that directly impact teachers' careers. Limitations of VAMs must be strongly considered in the design of teacher evaluation systems, and they should be used in conjunction with other measures for data triangulation (Braun, 2005; Loeb & Candelaria, 2012; Kane et al., 2013; Muñoz, Prather, & Stronge, 2011; Muñoz, Scoskie, & French, 2013; Stronge et al., 2007). Not surprisingly, these limitations also warrant careful consideration when designing VAMs for general classroom teachers who serve English learners; moreover, an additional set of variables unique to the EL population exists that compound these limitations for these teachers (Buzick, & Laitusis, 2010; Johnson, & Semmelroth, 2014; Jones, Buzick, & Turkan, 2013; Lakin, & Young, 2013).

For example, Gallagher (2002) was interested in the relationship between teacher evaluation scores and value-added measures on student achievement in a study conducted in a school where 85% of the population were English learners. Correlation analyses and HLMs were used to test this relationship, wherein the dependent measure was the state accountability reading test, the Stanford-9 [SAT-9]. In his methods section, Gallagher (2002) cautions, "all conclusions in this paper need to be understood within the limitations of the measurement instruments" (p. 9). That said, assessments administered to English learners that are designed for English speakers have been the subject of much debate in terms of validity (Abedi & Gándara, 2006; Hakuta & August, 1998; Jones, Buzick, & Turkan, 2013). The problem with using the SAT-9 in this study led to lack of correlation and measurement error, resulting in the likelihood of inaccurate results on student knowledge and skills. These errors have the potential to

negatively impact a teacher's VAM score as part of his or her evaluation (Jones, Buzick, & Turkan, 2013).

Using standardized tests as dependent measures of value-added models is highly problematic in terms of the EL population. Two important issues with testing concerns the non-standardized implementation of accommodations for ELs taking these tests and the notion that these students are taking a test in a language for which they are not fluent. These problems can lead to measurement error (Abedi, Hofstetter, & Lord, 2004). Accommodations need to be made because English learners may not be able to understand what the assessment is asking them to do in English although they may be able to perform the task in their native language.

Other related challenges with regard to VAM design involve the heterogeneity of English learners (Abedi & Gándara, 2006; Buzick, & Laitusis, 2010; Jones et al., 2013); for instance, testing accommodations for ELs have yet to be sufficiently and appropriately outlined by state education policy in terms of the students' educational history and their first language literacy skills (Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007), two features of heterogeneity among ELs. Second, recent immigrant English learners (as defined by NCLB, Title III) are exempt from reading achievement tests during their first 12 months in US schools, which excludes a population of newcomer students from value-added models. Measuring teacher effectiveness for this subgroup (within the EL subgroup) remains unknown. Moreover, more research is needed to assess differences in performance among heterogeneous subgroups of ELs, including long-term ELs and newcomers, as these two groups have unique contextual and instructional factors that may have a differential impact on performance (Turkan & Buzick, forthcoming).

Determining how to include EL demographic data in the development of a value-added model depends, in part, on the definition of EL (Loeb et al., forthcoming) and the student's unique factors (Turkan & Buzick, forthcoming). As per NCLB's Title III legislation, states are required to establish protocols for identifying and placing students in and exiting them from language development programs; the law also requires ELs to be on monitoring status for no less than two years after exiting the language program to ensure instructional supports are available if needed. As such, ELs are classified or reclassified with respect to participation in a language development program.

Toward that end, Loeb et al. (forthcoming) opted to include students identified as EL per administrative records collected from the school district participating in their study, as well as students who were classified as EL within the past three years, due to the complexity of identifying ELs from non-ELs. In doing so, they included students who may have recently exited the language development program and were on monitoring status for two years, possibly receiving language instructional support. This sampling approach exemplifies the types of decisions researchers, and potentially, practitioners, make when exploring the application of value added models to the EL subgroup.

Other factors to consider regarding value-added design is the possible influence of peer effects from non-ELs who were in the same classroom, and of other supporting teachers, such as ESL and Title I teachers, or instructional coaches, who may have added value unrelated to the classroom teacher (Jones et al., 2013). It is commonplace for ELs to receive a variety of instructional supports from a variety of specialists.

Considering the problems associated with VAMs in general, and for the EL subgroup, more research is needed to better understand their viability for measuring general classroom

teacher effectiveness related to ELs, especially as it relates to teacher evaluation systems. Jones et al. (2013) recommend that more research should be conducted on VAMs and the EL subgroup that explores variables associated with EL heterogeneity, and with issues regarding "inconsistent accommodation use across years…and changes in English proficiency or classifications of ELs" (p. 237).

Summary on inputs, processes, and outputs. Teacher qualifications and characteristics constitute the domain of inputs in the framework for teacher quality (Goe, 2007); inputs are what teachers bring with them to the classroom. Research findings in these two areas indicate inconsistency with regard to teacher qualifications such as certification and advanced degrees and effects on achievement for all students (Rowan, Correnti, & Miller, 2002; Sanders, Ashton, & Wright, 2005). Studies suggest secondary level math teacher credentials have positive effects on student achievement than other subject areas that have been researched (Goe, 2007). Findings from studies on teacher characteristics are also inconsistent with regard to ethnicity, race (Dee, 2004) and gender (Ehrenberg, Goldhaber, & Brewer, 1995). Less is known about EL teacher qualifications and characteristics although Masters et al. (2012) and Loeb et al. (forthcoming) have begun exploring these elements using teacher VAMs.

The second domain, teacher processes, includes school and classroom level practices. Much of the research that has been conducted in this area uses sophisticated statistical modeling to identify high performing and low performing teachers as measured by student achievement in order to explore the behaviors, attitudes, and beliefs of these groups of teachers with regard to best practices and quality teaching practices (Baker, Gersten, Haager, & Dingle, 2006; Muñoz, Prather, & Stronge, 2011; Strong, Gargani, & Hacifazlioğlu, 2011; Stronge, Ward, & Grant, 2011). Some studies use teacher perception surveys, observational data, and teacher evaluation

scores to predict student learning gains. Little, if any, research explores the domain of teacher processes in terms of the EL subgroup. Constructs of effective EL teacher practice center on principles of second language acquisition, and more specifically on academic language and literacy instruction (August & Shanahan, 2006; Baker, Gersten, Haager, & Dingle, 2006; Baker, et al. 2014; Calderón, Slavin, & Sánchez, 2011; Coady, Harper, & de Jong, 2011; Echevarria, Vogt, & Short, 2012; Gersten, Baker, Haager, & Graves, 2005; Lucas, Villegas, & Freedson-Gonzalez, 2008; Peercy, 2011; Téllez, & Waxman, 2005).

Teacher effectiveness is the third domain of the teacher quality framework (Goe, 2007) and is referred to as outcomes. This element strictly encompasses a teacher's contribution to gains in student achievement. Typically, VAMs are used to measure this contribution in both research and practice (Loeb & Candelaria, 2012), and while researchers support the use of VAMs to measure teacher effectiveness for several reasons, they also strongly caution their use in practice (Braun, 2005; Buzick, & Laitusis, 2010; Jones et al., 2013; Loeb & Candelaria, 2012). Careful consideration of variables used in building a value-added model is needed for teachers serving all children, including the English learner subgroup, as additional critical and unique factors are involved that can lead to statistical error if neglected, and subsequently, misinterpretation of teacher evaluation outcomes (Jones et al., 2013; Turkan & Buzick, forthcoming).

Teacher Evaluation Reform

The ESEA Flexibility Waiver and the Race to the Top grant program have incentivized State Education Agencies (SEAs) with waiving Adequate Yearly Progress (AYP) accountability targets and other mandates required by *NCLB* and with funding opportunities for turning around low-performing schools. In turn, SEAs must agree to reform their educator evaluation systems

to include multiple measures, such as: (a) data on student growth as measured by academic assessments, (b) standards-based observation rubrics, and (c) other acceptable measures, such as stakeholder surveys and teacher portfolios. From these combined measures, teachers earn a summative evaluation score that can be used to make personnel decisions (U.S. Department of Education, 2009). The majority of states have already transitioned to these new reform models, while some are undergoing this change initiative.

Considering the challenges associated with using value-added measures for evaluating general classroom teachers who serve English learners, teacher evaluation instruments, including observation rubrics, will play a prominent role in defining teacher quality among subgroups (Jones & Brownell, 2014). A notable difference between VAMs and teacher observation rubrics is that VAMs objectively and quantitatively provide the level of individual teacher effects on student achievement gains, whereas observation instruments provide opportunity and guidance for collecting observable evidence of classroom practices (Meyer, Cash, & Mashburn, 2011; Muñoz, Scoskie, & French, 2013). Moreover, evidence collected through observation instruments inform teacher growth and learning needs, provided the instruments are based on validated constructs of effective indicators of instruction, and evaluators have had sufficient training on how to use them (Danielson, 2012).

Standards-based Teacher Evaluation Instruments

Toward that end, a discussion of the literature on standards-based teacher evaluation instruments is warranted. Educator evaluation systems have undergone scrutiny in terms of the validity of measures used to assess teacher quality, and the outcome has resulted in national reform efforts to revamp those structures. For example, Weisberg et al. (2009) surveyed educator perception on teacher evaluation processes and found that practitioners regard the

process as a perfunctory compliance tool. Moreover, research in this area has found that many traditional forms of evaluation protocols are limited in their ability to render differentiation of teacher performance, precluding recognition of distinguished teachers and planning for needed growth and learning (Kane & Staiger, 2012; Sartain, Stoelinga, & Brown, 2011; Weisberg et al., 2009).

Nevertheless, researchers and practitioners alike use these standards based evaluation tools to gauge the quality of classroom instruction. For some researchers, the goal is to determine the extent to which observation evaluation scores influence student achievement (Goe, 2007; Jacob & Lefgren, 2005). They examine various observation frameworks to test their validity and reliability, which in turn informs stakeholders charged with recommending and oftentimes adapting selected frameworks to implement in policy and practice (Kane & Staiger, 2012; Kimball & Milanowski, 2009).

Several studies that use observation instruments in their designs suggest that these tools matter for student achievement (Baker et al., 2006; Gallagher, 2004; Gersten, Baker, Haager, & Graves, 2005; Heneman, Milanowski, Kimball, & Odden, 2006; Milanowski, 2004). Researchers have found that observation tools that use constructs of effective teaching indicators positively influence student achievement, suggesting instrument validity. As Meyer et al. (2011) argue:

> For observations to be useful for identifying effective teachers, improving classroom quality and teacher practice, or determining the features of student' classroom experiences that directly improve learning and development, observational measures must use standardized observation protocols that minimize measurement error and permit valid inferences. (p. 228–229)

However, most standards-based evaluation tools do not specifically measure teacher quality with regard to the instruction of English learners (Jones et al., 2013; Turkan et al., 2012). A discussion of widely used standards-based frameworks follows.

Teacher observation frameworks. A variety of observation instruments are used nationally, some of which include: (a) Charlotte Danielson's Framework for Teaching, (b) Stronge's Teacher Effectiveness Performance Evaluation System (TEPES), (c) Marzano's Causal Teacher Evaluation Model, (d) National Institute for Excellence in Teaching (NIET) TAP System Rubric, and (e) AdvancEd Effective Learning Environments Observation Tool (ELEOT). Many states and school districts have adopted these various tools and adapted them to fit their educational contexts. However, among these instruments, there is very little to no mention of how their indicators of instruction specifically support teaching and learning for the EL subgroup (Jones et al., 2013).

While some instruments address special populations, their indicators are vague at best (Jones et al., 2013). Instructional practices for ELs supported by robust research studies include the integration of academic language and content (August et al., 2009; Short et al., 2012), intensive vocabulary instruction (Carlo et al., 2004; Kieffer & Lesaux, 2012), opportunities for developing language and writing skills (Gomez, Parker, Lara-Alecio, & Gomez, 1996; Kim et al., 2011) and literacy practices that provide context-embedded strategies (August & Shanahan, 2006; Linan-Thompson et al., 2006). Indicators of instruction included in current models of teacher observation protocols lack these constructs of evidence-based practices for ELs. Jones et al. (2013) assert, "if instructional practices deemed effective for SWDs or ELs are not represented in the observation systems used by states/districts, it may provide disincentives for teachers to adopt such practices in their teaching" (p. 238).

The Danielson framework. Given its national recognition and widespread use, a discussion of research studies that center on the Danielson Framework follows. First, Kane, et al. (2012), Milanowski (2004), and Sartain, Stoelinga, and Brown, (2011) focused their studies on the Danielson Framework for Teaching (1996), an instrument designed to assist teachers and evaluators with the development and documentation of teacher practice. Findings from studies using the Danielson Framework indicate that the instrument demonstrates criterion-related validity (Kane et al., 2012; Milanowski, 2004; Sartain, Stoelinga, & Brown; 2011).

For example, in these studies, the observation instrument was positively related to student achievement. Sartain et al. (2011) found that teachers with high observation scores also had favorable VAM scores, while teachers with low observation scores had lower VAM scores, with students showing more and less growth in reading and math, respectively. These findings suggest that teacher evaluation scores are equivalent to effective teacher practices that impact student learning and thus render interpretations of the instruments as valid (Milanowski, 2004). Furthermore, researchers recommend that when using teacher observations as part of an evaluation system, teachers should be observed for more than one lesson with multiple, well-trained evaluators to increase reliability (Heneman, Milanowski, Kimball, & Odden, 2006; Kane et al., 2012).

The Danielson Framework (The Danielson Group, 2013) is composed of four domains of teaching: (1) Planning and Preparation, (2) Classroom Environment, (3) Instruction, and (4) Professional Responsibilities. Within these four domains, four rating scales are used to judge a teacher's level of effectiveness as unsatisfactory, basic, proficient, and distinguished. Each performance level includes a rubric containing critical defining attributes. Domains two and three apply directly to classroom observation while one and four entail the use of archival data

and other collectible evidence, such as lesson plans and recordkeeping (The Danielson Group, 2013). Indicators from all four domains are summarized in Table 4.

Table 4

Charlotte Danielson's Framework for Teaching Evaluation Instrument (The Danielson Group, 2013)

Domain 1 Planning and Preparation	1a Demonstrating Knowledge of Content and Pedagogy 1b Demonstrating Knowledge of Students* 1c Setting Instructional Outcomes 1d Demonstrating Knowledge of Resources 1e Designing Coherent Instruction 1f Designing Student Assessments
Domain 2 Classroom Environment	2a Creating an environment of respect and rapport 2b Establishing a culture for learning 2c Managing Classroom Procedures 2d Managing Student Behavior 2e Organizing Physical Space
Domain 3 Instruction	3a Communicating with Students 3b Using Questioning and Discussion Techniques 3c Engaging Students in Learning 3d Using Assessment in Instruction 3e Demonstrating Flexibility and Responsiveness
Domain 4 Professional Responsibilities	4a Reflecting on Teaching 4b Maintaining Accurate Records 4c Communicating with Families 4d Participating in the Professional Community 4e Growing and Developing Professionally 4f Showing Professionalism

* Mention of English Learners

The Danielson Framework for Teaching instrument is comprehensive with four domains and 22 indicators. Moreover, each indicator contains several features that provide more explicit details, as do the performance level criteria. While an expert forum on second language acquisition does not recommend expanding current observation instruments to include a comprehensive list of item indicators of EL teacher practices, it does recommend integrating critical indicators with existing models (National Comprehensive Center for Teacher Quality, 2012). Evidence-based constructs related to teaching practices for ELs included within such a framework would more precisely measure teacher performance with respect that subgroup.

Moreover, studies on teacher quality that use elements of the Danielson Framework as predictor variables on student achievement do not disaggregate subgroups in terms of dependent measures (Kane et al., 2012; Milanowski, 2004). What is known about the Danielson Framework is its tested validity with respect to students in the aggregate. Kane et al. (2012) suggest future research studies that explore subgroups in this context. Further, while recent findings (Loeb et al., forthcoming) support existing research that suggests good teachers tend to be good teachers for all students, they also find that instruction specifically related to teaching ELs improves EL student achievement in math. That said, all teachers and ELs would benefit from an instructional framework that includes these effective indicators (National Comprehensive Center for Teacher Quality, 2012).

Gaps in teacher evaluation models. Research on value-added models and teacher evaluation instruments lack variables in their designs essential to obtaining more precise outcomes of teacher quality regarding the EL subgroup. Given the issues related to VAMs and ELs, more emphasis will be placed on standards-based teacher evaluation instruments; however, most current standards-based models include little, if any, mention of quality teaching and

English learners (Jones et al., 2013; Turkan et al., 2012). Rather, evaluation instruments focus on general teacher processes (teacher practices) versus the unique practices for the EL subgroup.

Although the Danielson Framework for Teaching (2013) is comprehensive and has been vetted for its reliability and validity (Kane et al., 2012; Milanowski, 2004; Sartain, Stoelinga, & Brown; 2011), it does not shed light on general classroom teacher performance with respect to the instruction of English learners (Kane et al., 2012). Granted, the instrument was not designed to measure teacher performance with respect to specific subgroups; however, with the growing population of ELs and a widening achievement gap, it lacks critical features of effective instructional practices for ELs that school administrators and teachers need to know (Jones et al., 2013).

Although school level variables exist that have positive effects on outcomes for the EL population, such variables at the teacher level regarding this subgroup have been largely neglected in the literature on teacher quality (Loeb et al., forthcoming; Masters et al., 2012). For teacher observation instruments to render valid interpretations about EL teacher quality, research is needed to determine the extent to which teachers and their evaluators apply these instructional processes to practice (Jones et al., 2013); however, the instruments needed to measure the application of teaching practices for serving English learners are scarce.

Filling the gaps in teacher quality and evaluation models. Toward that end, Turkan et al. (2012) investigated these practices in their study addressing assessment of quality in the teaching of content to ELs. The purpose of their study was to define the knowledge base all teachers need to provide ELs with effective content instruction, including language arts, science, math, and social studies. The goal of this study was to contribute to a model for teacher licensure. To begin, they reviewed the literature on teacher standards to isolate specific knowledge and skills

content area teachers should possess, then they synthesized these knowledge and skills into two domains—pedagogical and linguistic. This process entailed (a) reviewing teacher state standards for teaching ELs, (b) constructing a framework of teacher knowledge and skills for teaching content to ELs through collaboration with researchers and practitioners, and (c) convening an advisory panel to validate these domains of practice.

These knowledge, skills, and attributes (KSA) were shared with an expert panel for review. An online survey regarding these specific items was disseminated to practitioners, researchers, and teacher educators to judge the importance of the KSA items gleaned from the standards and literature. The items were designed for assessing novice level teachers' knowledge and skills, as "there is a need for licensure tests designed to assess the targeted knowledge and skills…for all teachers exiting teacher education programs" (p. 5). Using a 5-point scale, survey participants were asked to rate the importance of each KSA statement, with 1 being *not at all important*, and 5 being *extremely important*. Once the online survey was completed, the ETS research team convened a 14-member panel to further review these statements and begin the construction of a teacher assessment.

The outcome of the ETS study (Turkan et al., 2012) resulted in a framework for general classroom teacher processes that exemplify important elements of practices for teaching content to ELs—a prototype measuring content teacher knowledge and skills in this specialized field. Given current teacher evaluation systems do not establish explicit expectations for how teachers should respond to unique needs of ELs, the extent to which teachers currently address them is unknown, and subsequently, less is known about teacher quality in light of ELs. The pedagogical and linguistic framework developed by Turkan et al. (2012) can serve as a tool to gauge general classroom teachers' current practice of developing their ELs' academic language

and content, both in the interest of research and practice. Research that explores teacher quality among general classroom teachers who serve English learners can illuminate the current state of practice and thus provide pathways for policymakers, researchers, and practitioners with an interest in improving education for all children.

Alabama is currently embarking in its educator evaluation reform initiative as per the ESEA Flexibility Waiver, and the opportunity to design a model inclusive of all learners presents itself to key stakeholders. Currently, Alabama policy allows for teachers who are certified in elementary education, foreign language, and secondary English as well as ESOL to be the ESOL teacher of record (Alabama State Department of Education, 2015). With baseline data indicating where general classroom teachers see themselves in terms of content and academic language instruction for ELs, research-based systems can be developed to support teacher and student growth in most pressing areas. With this goal in mind, the following four research questions examine constructs of teacher quality in Alabama related to its English learner population. Specifically, they address teacher qualifications, characteristics, and processes supported by empirical literature on effective teaching for ELs.

1. What number of factors in the *Teacher Quality* survey is identified through exploratory factor analysis?
2. To what extent do teachers report that they are prepared for teaching ELs academic language?
3. To what extent do teachers report they have adequate skills to teach ELs?
4. To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number

of ELs taught in the classroom predict teacher reported skill level for teaching academic language?

CHAPTER III. METHODS

Introduction

The Elementary and Secondary Education Act (ESEA) Flexibility Waiver allows state education agencies to opt out of accountability requirements introduced by *NCLB*; in return, SEAs must agree to reform their educator evaluation systems to include multiple measures that culminate in a summative evaluation score. These measures include: (a) evidence on student growth as measured by academic assessments, (b) standards-based observation rubrics, and (c) other measures that can include stakeholder surveys, teacher portfolios, and other data sources.

Two significant measures of teacher evaluation systems are standards-based teacher observation rubrics and student growth models, as they are the most widely implemented measures states and districts use (Jones, Buzick & Turkan, 2013). Evaluators who observe classroom instruction or who calculate a teacher's individual contributions to student growth when English learners are present should use protocols that incorporate factors unique to this subgroup to yield more valid and reliable teacher evaluation scores; however, current evaluation models lack these variables (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick & Turkan, 2013).

There was a need for this study, as gaps in the design of teacher evaluation systems exist with regard to teacher quality related to ELs, and thus examining the voices of general classroom teachers who work with English learners adds to the early discussion on this topic that can help fill this gap. In addition, the paucity of research on teacher quality in terms of serving ELs

during an era of teacher evaluation reform, compounded by the need to close the achievement gap between ELs and non-ELs, warranted the need for this study.

Finally, empirical studies suggest there are particular variables of interest related to teacher quality that predict differential effectiveness between the same general classroom teachers' ELs and non-ELs in mathematics (Loeb et al., forthcoming; Master et al., 2012). These variables are anchored in teacher inputs and were used to predict teacher effectiveness (outcomes). While Master et al. (2012) and Loeb et al. (forthcoming) were examining teachers who (a) share the same language as their ELs, (b) hold certification in bilingual/ESL education, and (c) indicate a preference for working in schools with high concentrations of ELs, among other input variables, my study extended their work by exploring the relationship these input variables have with teacher processes, the practices teachers use in their classrooms.

This chapter presents the research design for this study, including the sample, participants, survey instrument, data collection procedures, data analysis, and limitations. The purpose of this study was to examine two constructs of EL teacher quality, the inputs and processes as defined by Goe (2007), using quantitative methods. Specifically, this study surveyed general classroom teachers in Alabama schools regarding these two constructs to measure their reported skill level in developing ELs' academic language and content, preparedness in terms of training and certification, and characteristics with respect to teacher quality related to ELs. The following questions address this inquiry on teacher quality and English learners.

Research Questions

1. What number of factors in the *Teacher Quality* survey is identified through exploratory factor analysis?

2. To what extent do teachers report that they are prepared for teaching ELs academic language?

3. To what extent do teachers report they have adequate skills to teach ELs?

4. To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number of ELs taught in the classroom predict teacher reported skill level for teaching academic language?

Design of Study

I used a quantitative design in this study to examine descriptive statistics and relationships among variables which were calculated with data collected through a survey questionnaire on constructs of teacher quality related to ELs. Teachers in Alabama working in schools with approximately 20 or more ELs responded to the survey items. I began the study with an Exploratory Factor Analysis to determine if the Likert-type items in the survey instrument contained multiple factors measuring skills in teaching academic language. I used survey research to measure how general classroom teachers report their (a) level of preparedness to teach ELs; (b) skill level at teaching ELs academic language (c), and (d) characteristics, including preference teaching at schools with varied sizes of EL populations, as well as demographic variables such as race and age. Finally, I inquired on the extent to which these variables relate.

Sampling Procedures and Data Collection

I received permission from the Institutional Review Board (IRB) office to conduct the study (see Appendix A), which was located in the state of Alabama. Because the statewide EL population was approximately 3 percent at the time of my study, my sampling procedures were aimed at obtaining the largest sample of general classroom teachers possible to reach a maximum

sample size. In school districts in Alabama, the size of the EL population is not consistent, as they range from very small to large populations. Similarly, some districts may have small numbers, yet a particular school within that system may have a relatively large number of ELs. And by the same token, a district with large numbers of ELs may have schools that serve no ELs.

In light of these inconsistencies, I took the following steps for my sampling procedures: (a) identified and compiled a list of districts with 20 or more ELs, which corresponds to the state accountability sample size for Annual Measurable Objectives (AMOs); (b) cross-referenced these districts with data on school numbers to create a list of schools with 20 or more ELs; (c) from this list of schools, I contacted and provided school principals with an email that explained the research study and asked for participation of general classroom teachers in their schools; (d) attached the IRB approved consent information letter that contained the live link to the survey to the email to principals; (e) copied district superintendents and EL coordinators on this email letter; (f) sampled all general classroom teachers in schools identified as having 20 or more ELs during the current school year (2014–15).

I used an online software program called *Qualtrics* (www.qualtrics.com) to develop the quantitative survey. Participants received the survey in early February during the 2014–2015 school year. Once participants completed the electronic *Qualtrics* survey, I closed the survey in early April and downloaded data into the software application Statistical Package for Social Sciences (SPSS) for analysis.

Participants

While a total of 260 educators participated in the study, 223 surveys were completed. There was a total of 39 incomplete surveys; however, I filtered out 37 from the study because two contained insufficient information useful for descriptive analyses. In addition, of those 37, I

filtered out 14 surveys whereby participants were ineligible, as their reported current teaching assignment did not match the study's target participants. Data from a total of 223 participants were used in this study. Demographic data were collected via participants' self-reporting. Of the 223 participants, 29 (13%) did not respond to data collected on ethnicity nor age. Participant ethnicity consisted of Asian (0%), Hispanic (1%), African American (4%), White (80%), and Other (1% multi-racial and 1% Native American). Fifteen (7%) participants reported gender as male, and 179 (80%) as female.

Instrumentation

Teacher survey. The dependent variable used in the inferential statistical analyses was the construct of processes and practices that teachers use to develop ELs' academic content and language. The independent variables in this study were teacher qualifications and characteristics known as the inputs of teacher quality (Goe, 2007). A survey instrument was used to collect data on these variables. The survey prompts/stems used in the instrument were validated through an Educational Testing Service (ETS) study (Turkan, et al., 2012) that sought to determine the importance of specific knowledge, skills, and attributes that would be used to measure competencies of teachers of ELs in efforts to develop a teacher assessment for licensure. The Turkan et al. (2012) study describes how the survey was developed, validated, and tested for reliability. Upon review of state and national teaching standards for content-specific standards for teaching ELs, the ETS research team found "sweeping generalizations" (p. 4); thus, they conducted a literature review, validation survey, and convened a panel of practitioners and experts to determine which practices classroom teachers of ELs should use to develop academic language and content in order to measure their competency in knowledge, skill, and attributes.

In this ETS study (Turkan et al., 2012), sixty-seven statements on teacher quality for teaching academic language and content to ELs were developed under two domains: pedagogical and linguistic skills. Based on a literature review, these statements led to the construction of a framework of EL teacher knowledge. Within each domain (pedagogical and linguistic), general statements and statements specific to teachers of mathematics, English/language arts, science, and social studies were developed. To validate these statements, a national survey of practitioners and teacher educators was conducted to rate the level of importance of each item. The number of faculty in higher education sampled was 4000 with an extremely low response of only 64, and six thousand teachers were sampled with a total of 269 responses. Once the survey data were collected and analyzed, a 14-member panel served to further validate the statements "by reaching consensus that the statements support the claims of the assessment under development" (p. i).

For my research study, I used the prompts from the work of Turkan et al. (2012) but developed a different rating scale that addressed my research questions. As such, I adapted a survey for general classroom teachers using the statements validated by the ETS study (Turkan et al., 2012). I used the EL teacher practice statements developed through the ETS study as Likert-type items that measured teachers' reported level of skill. Teachers indicated their skill level using a five-point scale that used the following available responses "Less Skilled", "Somewhat Skilled", "Skilled", "Very Skilled", and "Highly Skilled". See Appendix 1 for the Teacher Survey.

The teacher survey also included questions pertaining to the construct of teacher inputs: the qualifications and characteristics (Goe, 2007), which were discussed in the literature review (Loeb et al., forthcoming; Master et al., 2012). I generated items for the same topics used in

these studies. The survey was used to address the three research questions and included 19 survey items that focus on two constructs of EL teacher quality—inputs and processes. Table 5 summarizes the items and their corresponding constructs.

Table 5

Summary of Teacher Survey Items and Corresponding Constructs

Survey Items	*Teacher Quality Constructs*
Part 1 1, 4–9 2–3 (demographic)	Inputs/Qualifications
Part 2: Pedagogical skills and practice for teaching ELs 10–12 (General)	Processes/Practices (Baker et al., 2014; Turkan et al., 2012)
Part 2 (cont.): Linguistic skills and practices for teaching ELs 13–14 (General)	Processes/Practices (Baker et al., 2014; Turkan et al., 2012)
Part 3 15–19 (Loeb et al., forthcoming; Master et al., 2012)	Inputs/Characteristics

Validity. "The extent to which our data-collection instruments, or processes, measure what they are supposed to measure is an indication of validity" (Ross & Shannon, 2011, p. 235). To measure the extent of validity of the survey I developed and adapted, I conducted a pilot

study of the teacher survey in early December 2014. A panel of teachers were selected to participate in the pilot study to determine if any weaknesses were present with the Teacher Survey. The survey items were based on constructs discussed in the literature review as well as the Turkan et al. (2012) study. Next, an Exploratory Factor Analysis (EFA) was run to determine the number of factors present in the Likert-type scale that support the purported constructs of these items. The EFA procedures are discussed later in Chapter 3, and results will be discussed in Chapter 4.

Teacher survey pilot study summary. Participant feedback from the pilot study centered on two areas: clarifying the wording of certain questions, and simplifying the appearance of the Likert-type survey items. Survey questions that participants deemed unclear included: (a) How many Continuing Education Units (CEU) hours of in-service training have you had in the last 5 years where the primary focus was on English learners; (b) What was the highest number of ELs in a class you taught last year? Using the drop-down menu, select the highest number at one point during that year; (c) What is the highest number of ELs in a class you currently teach? Using the drop-down menu, select the highest number at one point during this year? (d) If you have EL students in your classroom, or have had these students in recent years, do you speak the same language as any of your EL students? Given the number of similar comments participants made for each question listed, I revised all of these survey questions in the final survey for clarity and succinctness.

The second area for revision concerned the Likert-type scale questions. Comments regarding these questions included:

- *If there was any way to simplify items 9–16, I would*.
- *Way too loaded for one page*.

- *This page was a bit wordy. I would've probably been more willing to focus is [sic] there were only 3 on a page.*
- *A little too wordy. Simplification would be wonderful.*
- *Too many questions on one page.*

As a result of these comments, I included only 5–6 survey questions per page on the final electronic survey.

Participants made other, more isolated, yet pertinent suggestions. For example, one asked me to include Pre–K as a response to the question *What grade level do you currently teach?* I included this option given the possibility that some teachers participating in the survey may teach Pre–K. Another suggestion two participants made was to include the option *no preference* to the question *I prefer to work in a school that…(a) serves a large population of ELs; (b) serves a small population of ELs; or (c) has no EL population*. I reworded this question to provide a more positive tone and included the option *no preference*.

Reliability. The extent to which data-collection instruments "yield consistent results with minimal error is a demonstration of the concept of reliability" (Ross & Shannon, 2011, p. 235). To measure internal consistency of the surveys used in my study, I used SPSS to calculate a Cronbach's alpha index for each rating scale item (Ross & Shannon, 2011) measuring how teachers report skill level in developing English learners' academic language. An exploratory factor analysis was used to validate these 27 scale items, and reliability was tested using Cronbach's alpha. Reliability was extremely high ($\alpha = .986$), but acceptable as responses did not reflect a ceiling or floor effect. "Values around .8 are good" (Field, 2013, p. 715).

Data Analysis

I developed a data analysis plan to address each research question. Table 6 illustrates the alignment for each research question with its corresponding variables and analyses. Research question one was concerned with the number of factors in the *Teacher Quality Survey* as identified through exploratory factor analysis (EFA). Twenty-seven rating scale items were included in the survey to measure teacher reported skill level in teaching ELs academic language. With this number of rating scale items, EFA was used to determine whether any latent variables (Field, 2013) existed within the construct of teaching academic language. Exploratory factor analysis has "three main uses: (1) to understand the structure of a set of variables… (2) to construct a questionnaire to measure an underlying variable… and (3) to reduce a data set to a more manageable size while retaining as much of the original information as possible" (Field, 2013, p. 666).

Table 6

Summary of Research Questions, Variables and Analysis Tools

Research Questions	Variables	Analysis
1. What number of factors in the Teacher Quality survey is identified through exploratory factor analysis?	Practices for Teaching ELs Academic Language (27 Rating scale items)	Exploratory Factor Analysis
2. To what extent do teachers report that they are prepared for teaching ELs academic language?	Certification Teaching Assignment Grade Level Assignment Year Experience University Preparation In-service Preparation in EL Training Experience Teaching EL Pop by Size	Descriptive Statistics Descriptive Statistics (percentages, frequencies)
3. To what extent do teachers report they have adequate skills to teach ELs?	Practices for Teaching ELs Academic Language (27 Rating scale items)	Descriptive Statistics (percentages, means, sums, frequencies, and standard deviations)
4. To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number of ELs taught in the classroom predict teacher reported skill level for teaching academic language?	Variables a–d	Multiple Linear Regression, Hierarchical Multiple Regression

I used EFA to determine if the *Teacher Quality Survey* contained any latent variables that could be extracted as factors that would subsequently inform the number of dependent variables to examine statistically in terms of the relationships among teacher quality constructs.

Research questions two and three used descriptive analysis to explore the extent to which teachers report that they are prepared for teaching ELs academic language and have adequate

skills to teach ELs. To help answer research question two, respondents reported on their preparedness for teaching English learners academic language including (a) current certification area/s; (b) current teaching assignment; (c) grade level assignment; (d) years of teaching experience; (e) university coursework in ESL; (f) hours of in-service training, including Continuing Education Units (CEUs) for professional development on the instruction of ELs; and (c) the number ELs taught in the prior and current school year. Descriptive statistics were calculated to for each variable.

Research question three analyzed the extent teachers report skill level in pedagogical and linguistic practices for teaching English learners academic language using the rating scale. Descriptive statistics were run on these rating scale items to show percentages, means, and standard deviations for each of the 27 survey items.

Research question four used multiple regression to measure the extent to which teacher quality inputs (qualifications and characteristics) predict teacher skill level for teaching academic language (processes). Four independent categorical variables were used in this regression model: (a) certification, (b) preference working with EL populations, (c) in-service preparation in EL training, and (d) number of ELs taught in the classroom. Because the independent variables were categorical, dummy variables were used in SPSS to run the analysis. The dependent variable was the summed score for teacher reported skill level for teaching academic language to ELs; thus, a new variable was computed in SPSS to present that score. Incomplete responses were filtered out prior to calculating summed scores.

Summary

Chapter 3 discussed the study's methods, participants, instrumentation, data collection procedures, and an alignment plan used to map the statistical analyses with the four research questions. In addition, validity and reliability of the survey was described, both in terms of the pilot study and the exploratory factor analysis that was performed to determine the number of factors in the rating scale items included in the *Teacher Quality Survey*. Finally, data analysis procedures were discussed. Chapter 4 will present the results of the quantitative analyses that answer research questions one through four.

CHAPTER IV. RESULTS

As stated in Chapter 1, the purpose of this study was to explore relationships among teacher quality constructs (Goe, 2007) related to English learners. Namely, this study measured teacher qualifications, preparation (inputs), and instructional practices (processes) for teaching academic language to EL students in Alabama. The problem this study addressed lies in the prevalence of teacher quality and evaluation models that lack design constructs integral for teaching ELs (Calderón, Slavin & Sánchez, 2011; Jones et al., 2013), a fast growing subgroup of students in Alabama schools.

Chapter 4 begins with a discussion about how I handled missing data then provides results for research questions one through four. The Exploratory Factor Analysis answers research question one, and research question two explores descriptive statistics related to teacher qualifications and characteristics. Next, descriptive statistics pertaining to research question three concerning how teachers report skill level in teaching academic language and content to ELs are summarized. Finally, results from five multiple regression analyses are presented for research question four, the extent to which teacher qualifications and characteristics predict skill level on teaching ELs academic language. This chapter concludes with a summary and brief introduction to Chapter 5.

Missing Data

There were a total of 39 surveys with either missing data (skipped items) or that were incomplete. Sixteen participants with incomplete surveys completed all of the demographic

questions in part one of the survey; of those, eight completed only part of the rating scale items while eight did not complete any of those items. These sixteen participants did not complete part three which included the rest of the demographic survey items. Missing data can be identified as non-completion or general non-response, the latter of which is more prone to bias results (Culbertson, n.d.). The pattern observed here is more consistent with non-completion, which is similar to attrition in longitudinal studies where subjects decide not to complete the study mid-way. These 16 participants were omitted from all analyses.

To answer research questions two and three, I filtered out another fourteen participants' responses based on their current teaching position. The justification for filtering out the participant responses were (a) seven teacher participants were currently assigned to teach English as a Second Language (ESL) and one was a bilingual aid, so the number of ELs they taught for each three years (variables) were understandably very high, as that is the nature of an ESL teacher's teaching context; (b) two participants were currently assigned as administrators and do not teach in the classroom; (c) three participants' were currently assigned as physical education teachers who may teach combined classes which possibly accounts for the higher numbers of ELs they reported for these variables; and (d) one indicated certification in ESL, yet did not indicate the current position held; this participant also selected "not applicable" to the variable for numbers of ELs taught last year, and skipped the item for numbers taught this year.

Research Question 1: Exploratory Factor Analysis.

An Exploratory Factor Analysis (EFA) was used to determine how many factors were in the Teacher Quality survey and to test the feasibility of making the survey more manageable with fewer items. I expected two factors to emerge through the factors analysis based on the work of Turkan et al. (2012), which classified all stems as falling under the pedagogical or

linguistic domain. Prior to running an EFA several assumptions must be met. First, sample size affects the reliability of the factor analysis, a concept that involves the ratio of participants to variables. Debate exists among statisticians as to an acceptable benchmark, and Field (2013) posits that in general, sample size is most critical in satisfying these benchmarks. This study used the *Kaiser-Meyer-Olkin* (KMO) measure of sampling adequacy test, which "represents the ratio of the squared correlation between variables to the squared partial correlation between variables" (Field, 2013, p. 684). The KMO measure verified the sampling adequacy for the analysis, KMO = .977, well above the acceptable limit and falling in the range of "marvelous" (Field, 2013). Second, I used *Barlett's test of sphericity* ($p < .001$), to determine if the correlation matrix was significantly different from an identify matrix. This test usually results as significant, but in the event it does not, it can be problematic (Field, 2013).

A second assumption in EFA is normal distribution of data. Variables should be normally distributed at the interval level if the researcher intends to generalize findings or test for significance (Field, 2013). A benchmark for normality is skewness and kurtosis; the former may reveal data either have a negative or positive skew on a distribution curve, whereas kurtosis indicates if the data are either flat, or peaked and centered on the curve. All rating scale variables met the assumption for multivariate normality and will be reported in Chapter 4.

Factor Extraction

After screening data for sampling adequacy and correlation among variables, a principal axis factor analysis was conducted using SPSS version 22 with varimax rotation (orthogonal), a process that improves interpretability of factors (Field, 2013). An initial analysis was run to obtain eigenvalues for each factor in the data. "Eigenvalues associated with a variate indicate the substantive importance of that factor. Therefore it is logical to retain only factors with large eigenvalues" (Field, 2013, p. 677). In this analysis, one factor had eigenvalues over Kaiser's criterion of 1 and in combination explained 73.042% of the variance. The scree plot (Figure 2) clearly showed inflexions that justify retaining one factor. Although the adapted rating scale items measure knowledge, skill, and attributes for teaching English learners academic language (Turkan et al., 2012) using two domains—pedagogical and linguistic, only one factor emerged as a veritable construct of academic language instruction for ELs. Because only one factor was extracted, varimax rotation could not be performed. Reliability for the rating scale items was high (α = .986) but acceptable.

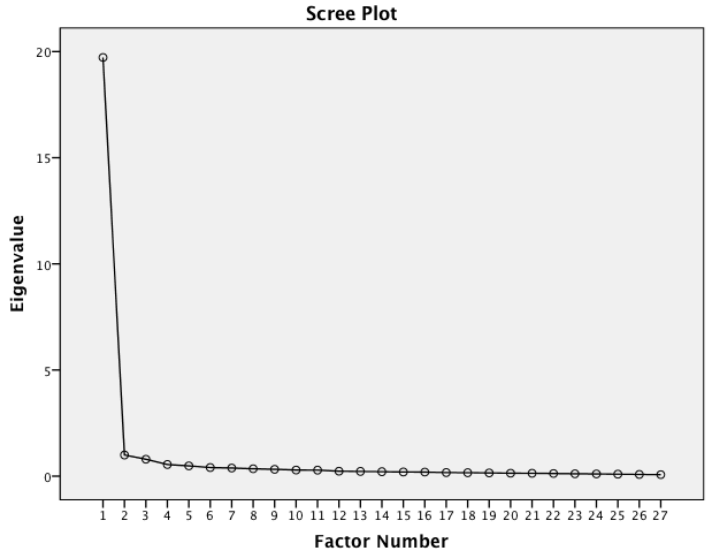

Figure 2. Exploratory Factor Analysis (EFA) Scree Plot

Research Question Two: **Teacher Reported Preparation to Teach ELs**

A construct of teacher quality related to ELs concerns teacher qualifications and preparedness, including certification, teaching and grade level assignment, years of experience, university preparation, in-service preparation in EL training, and experience teaching ELs in terms of numbers taught per year. To answer research question two, descriptive analysis results are reported below.

Certification, Teaching and Grade Level Assignment

Teachers reported their certifications, including 168 (75%) in elementary education, 37 (17%) in reading, 58 (26%) in early childhood education, 30 (14%) in math, 25 (11%) in science, 30 (14%) in social studies, 27 (12%) in English/Language Arts, 13 (6%) in English as a Second language (ESL/ESOL), and 26 (12%) in other. Other certifications included business marketing, marketing education, special education, library media, Spanish, and school counseling. Participants had the option to report more than one area of certification. The high percentage of teachers who reported having certification in elementary education reflects state data in that a) there are more elementary schools than secondary level schools, and b) the English learner population is concentrated at the elementary level.

Data for current teaching assignments included 159 (71%) in elementary education, 41 (18%) in reading, 48 (22%) in math, 38 (17%) in science, 43 (19%) in social studies, 34 (15%) in English/Language Arts, 7 (3%) in elective classes, and14 (6%) in other. Teaching assignment listed as other included alternative school, health, library media, co-teacher (special education), spelling, intervention teacher, and special education. Table 7 presents data related to grade level assignment both in terms of grade span groups and for grades pre–K through 12. Again, the majority of participants reported their teaching assignment as elementary education (71%). The

percentage of teachers who reported content area teaching assignment was slightly higher than reported certification in the content area.

Table 7

Current Grade Level Assignment as Reported by Survey Participants

Primary 59%		Intermediate 40%		Middle 14%		High 14%	
PK	4 (2%)	Gr 3	38 (17%)	Gr 6	8 (4%)	Gr 9	6 (3%)
K	43 (19%)	Gr 4	31 (14%)	Gr 7	12 (5%)	Gr 10	8 (4%)
Gr 1	44 (20%)	Gr 5	20 (9%)	Gr 8	10 (5%)	Gr 11	8 (4%)
Gr 2	40 (18%)					Gr 12	7 (3%)

Years of Experience, Advanced Degree, Training, and Experience with ELs

Teacher reported years of teaching experience ranged from 1 to 44 years with median at 10. In terms of advanced degrees, 12 (5%) teachers reported earning a master's degree in ESL/TESOL, 110 (49%) had a master's degree in education, 14 (6%) had an Education Specialist's degree, and one person reported earning a doctoral degree. Twenty-six (12%) teachers reported *other* including a master's degree in special education, master's in counseling, master's in instructional leadership, and doctoral degree in leadership.

Of the 211 (95%) teachers who reported they did not have an advanced degree in ESL/ESOL, 48 (22%) reported taking university level coursework in ESL/ESOL. Further, teacher participants reported on the number of professional development (PD) hours earned, including Continuing Education Units (CEUs) related to the instruction of ELs. Twenty-three (10%) reported having zero hours of PD, 108 (48%) reported having 1–10 hours, 34 (15%)

reported 11–20 hours, 29 (13%) reported 21–30 hours, 2 (1%) reported 31–40 hours, and 15 (7%) reported earning more than 40 hours of EL training. Thus, over half (58%) of the respondents have had little training related to English learners, while 8% reported having substantial (more than 31 hours) of EL-related training.

Finally, Master et al. (2012) suggested that teachers who have prior experience teaching about six EL students or more at a time in the classroom have a positive impact on EL student outcomes. Thus, data were collected on the largest number of ELs taught during the current school year and last school year. While these data are used to answer research question four, they also address research question two in terms of teacher preparation. Results for both years are very similar with approximately 15% of teachers reporting three to four as their largest numbers of EL students taught and approximately 36% of teachers reported five to nine as their largest number taught. Lastly, 11% of the participants reported having more than 10 ELs as the highest number of ELs taught over the course of both school years. In sum, nearly a majority (47%) of respondents reported having at least five EL students in both years.

Research Question 3: Teacher Reported Skill Level to Teach ELs

To what extent do teachers report they have adequate skills to teach ELs? This study focused on constructs of teacher quality for ELs based on teacher knowledge and skills for developing academic language for ELs (Turkan, et al., 2012), a construct with a research base with strong evidence for improving literacy and academic success (Baker et al., 2014). To answer research question three, I used a five-point rating scale to measure teacher-reported skill level in the development of academic language for ELs. Twenty-seven items compose the rating scale. Descriptive results for the three highest and lowest means from the rating scale items are

discussed below. Table 8 summarizes the results (see Appendix 2 for a complete summary of descriptive data). Normality for all rating scale data was met.

Table 8

Three Highest and Lowest Means for Teacher Reported Skill Levels

	Mean SD	Frequencies				
		Less Skilled No. (%)	*Somewhat Skilled* No. (%)	*Skilled* No. (%)	*Very Skilled* No. (%)	*Highly Skilled* No. (%)
Highest 3 means						
a. Provide ELs with oral and visual support (e.g., think-alouds, word picture cards) to produce language orally and in writing.	3.270 0.995	4 (1.8)	44 (19.7)	70 (31.4)	59 (26.5)	23 (10.3)
b. Teach ELs new vocabulary in context, both intentionally and incidentally.	3.000 1.035	10 (4.5)	59 (26.5)	71 (31.8)	42 (18.8)	18 (8.1)
c. Provide multiple opportunities for ELs to process content in group contexts, including with their English-speaking monolingual or proficient bilingual peers (e.g., peer work assignments, whole-class discussions).	2.990 1.042	11 (4.9)	61 (27.4)	65 (29.1)	48 (21.5)	16 (7.2)
Lowest 3 means						
d. Build on ELs' knowledge of cognates between English and their home language.	2.240 1.041	52 (23.3)	73 (32.7)	46 (20.6)	17 (7.6)	6 (2.7)
e. Know that the discourse of academic texts in content areas includes the use of passive voice in describing events or explaining cause and effect.	2.360 0.973	36 (16.1)	81 (36.3)	52 (23.3)	21 (9.4)	4 (1.8)
f. Encourage critical thinking by acknowledging ELs' diverse cultural experiences (e.g., arguments from authority may be highly valued in some cultures).	2.390 1.009	39 (17.5)	75 (33.6)	53 (23.8)	25 (11.2)	4 (1.8)

In short, teachers reported highest and lowest skill levels with regard to six particular areas of teaching EL academic language. First, participants reported adequate to high skill level for providing oral and visual support for productive (written/spoken) skills (37% of teachers),

vocabulary in context (meaning-based) (27% of teachers), and strategic grouping configurations for ELs (28% of teachers). Conversely, the areas where teachers reported having lower skills were in developing cognate knowledge (56% of teachers), discourse of academic texts which uses passive voice to explain cause and effect (52% of teachers), and encouraging critical thinking and cultural diversity (52% of teachers). These data combine the two lowest and two highest rating scales (less skilled with somewhat skilled and very skilled with highly skilled), respectively.

To summarize, over half of the teachers reported feeling low to somewhat skilled in these three particular areas of knowledge and skills for teaching ELs academic language, while a considerably smaller percentage (less than 37%) of teachers reported very skilled to highly skilled. Notably, Turkan et al. (2012) identified these items as either pedagogical or linguistic skill domains. The top three means all fell under pedagogical, whereas the bottom two of three means fell under linguistic. Although the third lowest was classified under pedagogical, the item focused on the notion of culture as it plays into critical thinking. This particular skill as it relates to teaching EL students academic language may overlap with the linguistic domain given its theoretical basis in second language acquisition and bilingualism (Baker, 2006). That said, teachers reported having lower skills in the area of linguistic versus pedagogical instruction, which may have practical implications.

Research Question Four: Relationships Among Teacher Quality Variables

To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number of ELs taught in the classroom predict teacher reported skill level for teaching academic language? To answer question four, some survey questions were modified or combined. The first independent

variable, *Preference Working with the EL Population*, was presented as a survey item with four options: large, small, none, and no preference. I transformed this variable into a different variable (*Preference Teaching EL Population Sizes*) with three levels: large, small-to-none, and no preference (I simply combined small and none because they are essentially the same question). Next, I transformed current grade level assignment into a different independent variable with four levels: primary, intermediate, middle, and high. In addition, I computed two new variables. For the first, I combined for a summed total two consecutive years of largest numbers of ELs taught in the classroom; for the second, I used all 27 rating scale items to calculate a summed total rating score for the dependent variable as results from the exploratory factor analysis indicated only one factor was in the rating scale items. For all categorical variables (e.g., preference working with population sizes of ELs, in-service preparation hours in EL training, teacher certification, coursework in ESL/ESOL, grade span assignment), I created dummy variables prior to setting up the analysis. One variable, *the number of ELs taught in the classroom* was continuous.

Multiple Linear Regression

Assumptions of the Linear Model

Prior to running a multiple linear regression (MLR) analysis, assumptions must be tested. They include independent and random sampling, additivity and linearity, independent errors, bivariately normally distributed variables, homoscedasticity, and multicollinearity (Fields, 2013). As stated in Chapter 1 under limitations, random sampling did not occur due to the nature of the topic at hand. For example, not all schools in Alabama have a sufficient number of ELs enrolled, rendering the *Teacher Quality Survey* irrelevant to potential participants. The target sample was general classroom teachers serving in schools with roughly 20 or more ELs, and my aim was to

gain an acceptable sample size for the survey, which precluded random sampling. Additionally, because the survey was web-based and disseminated statewide, I was not able to ensure teachers completed the survey independently.

Linearity and homoscedasticity are tested simultaneously through scatterplots of the values of the residuals against the values of the dependent variable. Residuals are the errors depicting the predicted versus the actual score from the dependent variable. "If linearity and homoscedasticity hold true then there should be no systematic relationship between the errors in the model and what the model predicts" (Fields, 2013, p. 192). Toward those ends, the graphic representation of the standardized predicted values and standardized residual values (*zpred vs*. *zresid*) should appear as a "random array of dots" (Fields, 2013, p. 350). For all five of the regression models I analyzed, these two assumptions were met and are discussed in their respective sections below.

To test the assumption for multivariate normal distribution, the researcher must determine if the dependent variable is normally distributed in the population for each level of the independent variables (Ross & Shannon, 2008). This test can be analyzed using a histogram, looking for skewness and kurtosis, and with a normal probability plot, whereby data should lie on or very near the diagonal (Fields, 2013). The assumption for multivariate normal distribution was met for each multiple regression analysis and discussed in its respective section.

Next, it is essential to determine independent errors. "For any two observations, the residual terms should be uncorrelated (i.e., independent)…if we violate the assumption of independence, then our confidence intervals and significance tests will be invalid" (Fields, 2013, p. 311). For each analysis I tested the assumption using the Durbin-Watson test and all met the assumption. The optimal statistic to look for is as close to 2 as possible.

To assess the assumption of multicollinearity, the researcher must determine whether predictors in the analysis are collinear, a potential problem that occurs when two or more predictors are included in the regression analysis (Ross & Shannon, 2011). If independent variables are too highly correlated (e.g., above .80 or .90), "it becomes impossible to obtain unique estimates of the regression coefficients because there are an infinite number of combinations of coefficients that would work equally well" (Fields, 2013, p. 324). To check for multicollinearity, a variance inflation factor (VIF) above 10 is cause for concern; additionally, the tolerance statistic, its inverse, should be above 0.1 (Ross & Shannon, 2011). All regression analyses I ran in this study contained results that met the assumption for multicollinearity.

Lastly, to check residuals for evidence of bias, the researcher should investigate casewise diagnostics. "In an ordinary sample, we would expect 95% of the cases to have standardized residuals within about ±2" (Fields, 2013, p. 345). Cases with standardized residuals greater than 3 are cause for concern. In these MLR analyses, casewise diagnostics from SPSS indicated all standardized residuals were below 3, except for one which is discussed in its respective section.

Multiple Regression Analyses

Variables were explored using four multiple linear regression analyses followed by a hierarchical regression analysis using significant predictors determined by the previous four analyses. For the first set of four MLR analyses, the variables included certification, preference teaching size of EL population, hours of EL professional development, and the number of ELs taught in the classroom. To run the analyses, I created dummy codes for the categorical variables of interest, including all but *the number of ELs taught in the classroom*, a continuous variable. When running a multiple regression analysis using dummy variables, a baseline variable must be omitted from coding scheme for each independent variable, as it serves as the

referent from which to compare other related variables and can be in some instances be considered the control (Field, 2013). The referent can be selected either based on theoretical significance or if the variable consisted of a high mean, whichever is most meaningful for the analysis.

First MLR analysis: Certification. Once these variables were created, the first multiple linear regression model focused on certification and was analyzed using SPSS. The reference variable used was *certification in elementary education*, given it had the highest mean (75%), and I compared certification in reading, early childhood, math, science, social studies, English/Language Arts, and ESL/ESOL to it. Results are described below.

Assumptions. The histogram of residuals indicated a very slight positive skew, but overall, the data are symmetrical and bell-curved. The P-Plot showed a near fit on the line. The scatterplot did appear heteroscedastic, although in general, due to the nature of categorical variables (dummy variables), this test is more challenging to meet. The Durbin-Watson test resulted in 1.810, very close to 2, and casewise diagnostics showed only 2% of the cases listed, all of which were under 3. In terms of multicollinearity, all VIF statistics were well under 10, and tolerance well above .1. These data appear to have met the assumptions for MLR.

Results. The ANOVA revealed that the model was statistically significant ($p < .001$). R was .554, R^2 .306, adjusted R^2 .281, $F(7, 187) = 11.805$. The predictor accounts for approximately 30% of variation in teacher reported skill level in academic language instruction for ELs.

Clearly, two significant certification areas that predict teacher reported skill level are early childhood education and English as a Second Language (ESL/ESOL) ($p \leq .001$), respectively. In addition, both share a positive significant correlation with the outcome variable,

with ESL/ESOL at a moderate level $r = .509$ and early childhood relatively low, $r = .162$. Moreover, b-values (b) indicate that compared to teachers who are certified in elementary school, early childhood certified teachers tend to rate themselves 11 points higher on the rating scale, and teachers with additional ESL/ESOL certification tend to rate themselves 41 points higher, an unequivocal difference. All other variables resulted in a non-significant relationship. However, all content-based certification (math, science, and social studies) indicated a negative correlation, yet English/Language Arts revealed a positive, yet very low correlation $r = .010$. When taken alone, without other possible predictors of reported skill level, certification in ESL/ESOL and early childhood education matter more than certification in elementary education.

Second MLR analysis: Preference teaching EL population by size. The second multiple linear regression model explored preference teaching EL population by size and the rating scale outcome variable. This time, the reference variable was *preference for teaching a large EL population*, with the following comparisons variables: (a) preference for teaching small-to-no ELs, and (b) no preference. I selected the reference variable under the empirical basis that it suggests a positive relationship with student learning outcomes (Masters et al., 2012).

Assumptions. The histogram revealed a near symmetrical bell curve, and P-P Plot indicated data fit the line although with slight sag. The scatterplot of residuals indicated vertical lines forming across the graph; categorical variables may contribute to these results. Durbin-Watson was 1.684, close to the benchmark of 2. Checking for multicollinearity, VIFs were both 1.604, and tolerance levels were all higher than .624. Casewise diagnostics indicated

standardized residuals were below 3, with only 4% of the cases listed. Thus, 96% of the cases were sufficiently within the boundaries of ±2.

***Results*.** The ANOVA indicated this model accounted for significant variance ($p \leq .001$). R was .378 indicating a moderate multiple correlation, and R^2 was .143, adjusted R^2 .133. $F(2, 177) = 14.730$, $p \leq 0.01$. This predictor accounts for approximately 14% of variation in teacher reported skill level in academic language instruction for ELs. Both predictors significantly predict the rating scale outcome variable ($p \leq .001$). Negative correlations exist between the outcome variable and qualitative variables, indicating that the reference group (prefer large EL population) had the highest means. To illustrate, those with a preference for teaching small to no EL population show a negative correlation $r = -.226$, $p = .001$, and those who indicated no preference have a negative correlation $r = -.100$, $p = .091$. Coefficients (b) indicate that compared to teachers who prefer to work with large populations of ELs, those who prefer teaching small to none rate themselves 23 points lower on the EL instructional skills rating scale, and those who reported no preference rate themselves 17 points less. In sum, while having no preference with regard to teaching EL population size is associated with somewhat lower ratings of skills, preference for teaching a small to no EL population is associated with much lower ratings of skills.

Third MLR analysis: EL professional development (PD) hours. I ran a third MLR analysis to explore the predictor variable for hours of EL professional development. The referent variable was *more than 40 hours of EL professional development*, given the theoretical and empirical basis that more intensive training and preparation increases knowledge and skills in teaching EL students academic language (Coady et al., 2011; Masters et al., 2012).

Assumptions. The histogram for standardized residuals shows good symmetry, and the P-P Plot indicates data fit the line. The Durban-Watson test was 1.799, very close to the benchmark of 2. In checking for multicollinearity, VIFs were under 3.694, and the lowest tolerance level was .271, with the remaining tolerance statistics over 4.02. Casewise diagnostics indicated standardized residuals were below 3, except for one case, which was 3.031. Only 3% of the cases were listed. Thus, 97% of the cases were sufficiently within the boundaries of ±2.

Results. The ANOVA for the model indicated a good overall fit ($p \leq .001$). R was .514 indicating a moderately strong multiple correlation, and R^2 was .264, adjusted R^2 .243. $F(5, 174) = 12.491$, $p \leq 0.01$. Together the predictors account for approximately 26% of variation in teacher reported skill level in academic language instruction for ELs. All but one predictor significantly predicted the rating scale outcome variable. Negative correlations exist between the outcome variable and two categorical variables, indicating lower knowledge ratings compared to the reference group (40 plus hours of EL PD). Positive low correlations exist between the outcome variable and Twenty One to Thirty Hours EL PD $r = .223$, $p = .001$. The other two positive correlations were not statistically significant compared to the reference group.

Coefficients (b) indicate that compared to teachers who reported having more than 40 hours of EL professional development, those who reported having (a) zero hours of EL PD rate themselves 45 points less $p \leq .001$, (b) one to ten hours of EL PD rate themselves 33 points less $p \leq .001$, (c) eleven to twenty hours of EL PD rate themselves 24 points less, $p \leq .001$, and (d) twenty one to thirty hours of EL PD 15 points less $p = .025$. The last variable (thirty one to forty hours of EL PD) was not significant but indicated that teachers tend to report skill level 6 points less. In short, teachers who reported having zero to ten hours of EL professional development tend to report a much lower practice and skill level than those who report having at least 30 or

more hours of professional development related to teaching English learners. A clear linear association exists between professional development hours and self-reported knowledge and skills, as well as a 27-point gap in knowledge and skills between these two groups.

Fourth MLR analysis: Number of ELs taught in the classroom. This analysis used a continuous independent variable; therefore, no referent variable was needed to build the model. I explored the how teachers reported the number of ELs taught in the classroom this school year and last school year as predictors of teacher reported skill level in teaching academic language to ELs, as these predictors were found in prior empirical research to have a significant relationship with student learning outcomes (Master et al., 2012).

***Assumptions*.** The histogram of standardized residuals indicates a symmetrical bell curve, and the P-P Plot of standardized residuals depicts a well-fitted line. The scatterplot of standardized residuals against standardized predicted values suggests linearity and homoscedasticity, as the values appear as a "random array of dots" (Fields, 2013, p. 350). The assumption for independent errors was met with a Durbin-Watson test at 1.549, and multicollinearity was also met as both tolerance statistics were at 3.78, and VIF for both were at 2.643. Casewise diagnostics showed all standardized residuals were below 3, with only 4% of the cases listed. Ninety-six percent of the cases were within approximately ±2. Toward those ends, assumptions were met.

***Results*.** The ANOVA for the model indicated a good overall fit ($p \leq .001$). R was .291 indicating a moderate to low multiple correlation, and R^2 was .085, adjusted $R^2 = .074$. $F(2, 175) = 8.096$, $p \leq 0.01$. Together the predictors account for approximately 8% of variation in teacher reported skill level in academic language instruction for ELs. One predictor, Number of ELs Taught Last School Year, significantly predicted the rating scale outcome variable ß = .247, $p =$

.037. Positive correlations exist between the outcome variable and two categorical variables: the number of ELs taught last school year $r = .289$, $p \leq .001$, and number of ELs taught this school year $r = .248$, $p \leq .001$, both relatively strong.

The ß-value indicates that as the reported number of ELs taught last school year increases, teacher reported skill level in teaching academic language to ELs increases by 1.683, $p = .037$, and as the reported number of ELs taught this school year increases, reported skill level increases by only .385, a marked difference, although the latter was not statistically significant. Moreover, standardized beta values (ß) "are all measured in standardized deviation units and so are directly comparable: therefore, they provide a better insight into the 'importance' of a predictor model" (Field, 2013, p. 340). So, for every increase in the reported number of ELs taught last school year, the reported skill level increases by .247. Hence, the larger number of EL students taught in the prior school year had a significant, positive effect on teacher reported skill level, although causation was not known.

Hierarchical Multiple Regression (HMR)

A HMR analysis was used to address research question four which examines the extent to which the combined independent variables (a) certification, (b) preference working with population sizes of ELs, (c) in-service preparation hours in EL training, and (d) the number of ELs taught in the classroom together predict the dependent variable, teacher reported skill level for teaching academic language to English learners as operationalized by the *Teacher Quality Survey* rating scale items (Turkan et al., 2012). In this HMR analysis, I entered each independent variable in blocks in the order of importance based on prior research (method suggested by Fields, 2013). The order in which I entered blocks of independent variables in the hierarchical

multiple regression model follows: (a) Certification, (b) Hours of EL Professional Development, (c) Preference Teaching EL Population by Size, and (d) Number of ELs Taught.

Assumptions. Assumptions for hierarchical multiple regression are the same for multiple linear regression (Fields, 2013). Benchmarks for meeting multivariate assumptions were discussed at the beginning of Chapter 4. Next, I discuss how assumptions were met for the final HMR model. To test for linearity and homoscedasticity, these two assumptions were met as illustrated in Figure 3, as the scatterplot shows random dots across the graph. To test the assumption for multivariate normal distribution, I used a histogram to analyze the data looking for skewness and kurtosis, and I used a normal probability plot to determine if the data lie on or very near the diagonal (Fields, 2013). The histogram in Figure 4 shows a well-shaped distribution where data in this multiple regression study are nearly symmetrical and approximately bell-shaped, and the P-P Plot in Figure 5 indicates these data lie very close on line. The assumption for multivariate normal distribution was met.

In determining independent errors, I used the Durbin-Watson test, which resulted in a level of 1.767, very close to the optimal level of 2. In testing for multicollinearity VIF levels for all for blocks were all below 5.148 and the tolerance statistic for each model were well above .194, the lowest level. Finally, casewise diagnostics were all under 3 (case number 123 at 2.994) with only 3% of cases listed.

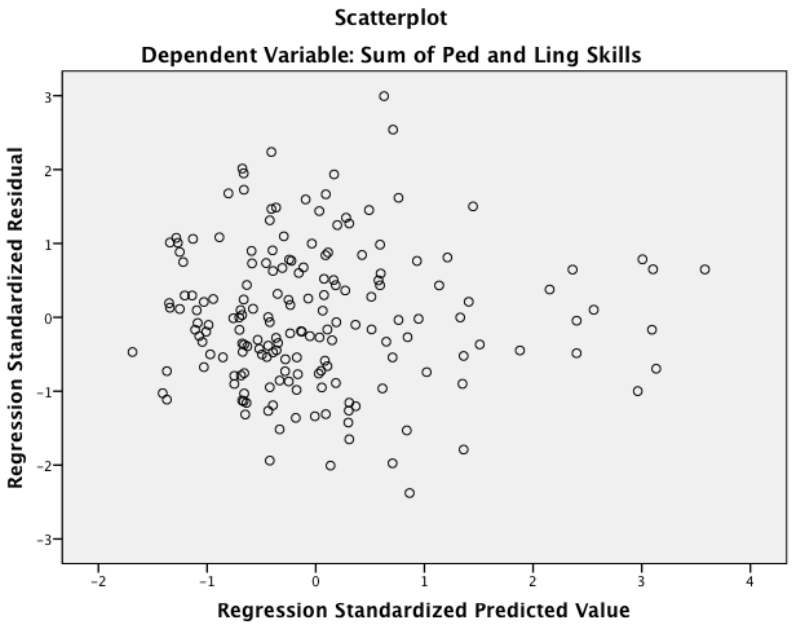

Figure 3. Scatterplot of zpred. vs. zresid. Scores

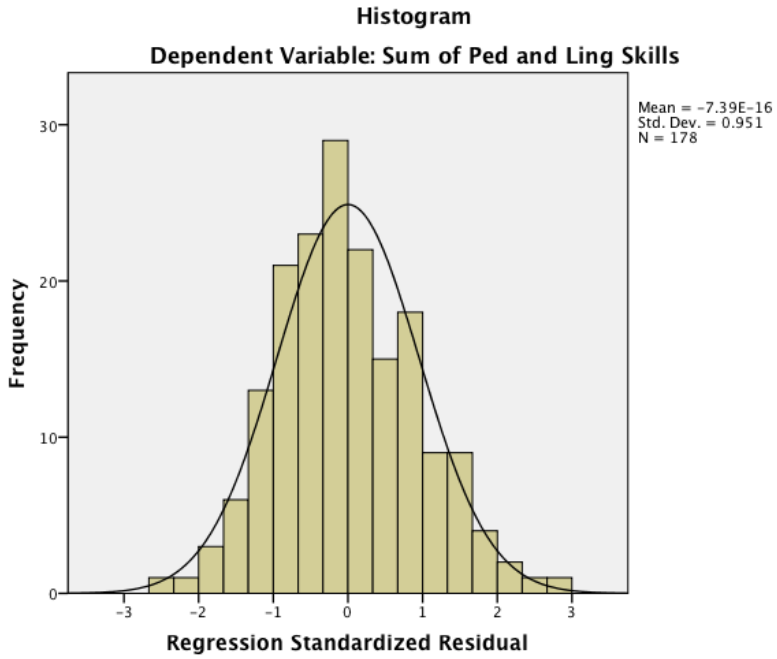

Figure 4. Residuals for Sum of Pedagogical and Linguistic Skills

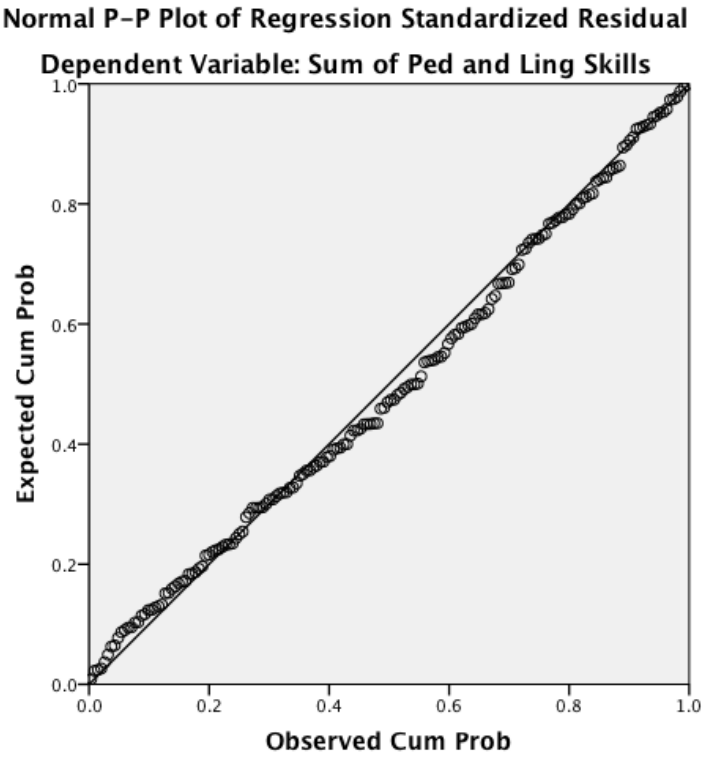

Figure 5. Observed Cumulative Distribution Against Predicted Distribution

Results. ANOVA for all four models revealed that the hierarchical multiple regression analysis significantly improved my ability to predict teacher reported skill level teaching EL students academic language ($p < .001$). Table 9 summarizes results of the final model with significant variables. Certification was entered first as the most important variable and used as the control. R^2 Change was .277, $p < .001$. However, when the other three predictors were added to the second block, Hours of EL Professional Development, R^2 increased to .395, R^2 change was .118, $p < .001$; when adding the third block of predictors, Preference Teaching EL Population by Size, R^2 was .421, R^2 change was .026, $p = .029$; finally, when adding the fourth block, Number of ELs Taught, R^2 was .422, R^2 change was .002, $p > .812$. These results indicate that all independent variables but the last, number of EL students taught, were significant in this model and explain a considerable amount of the variation in teacher reported skill level.

Table 9

Hierarchical Multiple Regression Results—Final Model with Significant Variables

Variable added to model	R	R^2	R^2 Change	Unstandardized Coefficients		Standardized Coefficients	
				B	Std. Error	Beta	Sig.
Certification	.526	.277	.277				
English as a Second Language (ESL/ESOL)				26.221	6.006	.301	<.001
Hours of EL Professional Development (relative to More than 40 Hours EL PD)	.628	.395	.118				
Zero Hours				-28.103	7.156	-.383	<.001
One to Ten Hours				-18.583	5.938	-.410	.002
Preference Teaching EL Population by Size (relative to Prefer Large Population)	.649	.421	.026				
Prefer Small to No EL Pop.				-11.081	4.132	-.225	.008

Model parameters are revealed in Table 9, as the *b*-values explain the relationship between the outcome variable and its significant predictors. Positive and negative values indicate respective direction of the linear relationship. There is a positive relationship between respondents who indicated having certification English as a Second Language (ESL/ESOL) and their reported skills in EL academic language instruction. So for teachers who report having certification in English as a Second Language (ESL/ESOL), the reported skills tended to increase by 26, $p < .001$. Clearly, having certification in ESL/ESOL matters to how teachers report knowledge and skills level for teaching academic language to English learners. Interestingly, having certification in Early Childhood Education when the second block was entered was at the significant level, $p < .05$ indicating a tendency for teachers to report slightly higher levels of skill.

Regression (b) values for the second and third blocks have negative relationships with the outcome variable when compared to the referent variables. For example, relative to teachers who reported having more than 40 hours of EL professional development, $p < .001$: (a) teachers who reported having zero hours of EL professional development reported 27 points less in teaching skills, $p < .001$, and (b) teachers who reported having one to ten hours of EL professional development reported 18 points less in teaching skills, $p = .002$. Finally, the third model suggests that when compared to preference for teaching a large population of EL students, $p < .001$, teachers who reported preferring to teach a small to no EL student population tend to report 11 points less in teaching skills, $p = .008$.

Based on this order of variables, certification is really important; professional development add significant variance beyond certification, and preference for teaching EL population based on size is still significant and important, controlling for certification and professional development. Controlling for all of these variables, number of ELs taught during the school year is not important in predicting self-reported knowledge.

Summary

The purpose of this study was to explore descriptively how teachers report their qualifications, characteristics, and skill level teaching academic language to English learners in Alabama and to examine the relationship among these teacher quality constructs using inferential statistics. Multiple regression analyses were used to test these relationships. Chapter 4 summarized the results of these analyses for the last three research questions, while the first addressed exploratory factor analysis. Chapter 5 will provide a brief synopsis of the significance and purpose of the study followed by the interpretation of results and their implications. In closing, Chapter 5 will offer recommendations for future research.

CHAPTER V. DISCUSSION

Summary

Currently, a plethora of teacher quality studies examine constructs of teacher quality in terms of teacher effectiveness and student outcomes (Goe, 2007; Kane & Staiger, 2013; Munoz, Prather & Stronge, 2011; Strong, Gargani, & Hacifazlioğlu, 2011; Stronge, Ward, & Grant, 2011; Stronge, Ward, Tucker & Hindman, 2007), as researchers are apt to do in today's era of teacher accountability and evaluation reform. But as we look at the whole picture of teacher quality, inclusive of English learners, there is a glaring gap in the area of practice (and processes) that warrants exploration. This gap is evident in teacher evaluation designs (Calderón, Slavin & Sánchez, 2011; Jones et al., 2013; Turkan, Croft, Bicknell, & Barnes, 2012), and in how we define teacher quality (Lucas, Villegas, Freedson-Gonzalez, 2008). This study veered in the direction of teacher quality for this special population. Specifically, this research study explored teacher qualifications and characteristics as functions of inputs that feed into the practices teachers use in the classroom and school context (processes) (Goe, 2007). Namely, these practices are the knowledge and skills teachers employ when teaching academic language and content to English learners (Turkan et al., 2012).

Survey Instrument

Research question one. *What number of factors in the Teacher Quality Survey is identified through exploratory factor analysis?* Exploratory factor analysis (EFA) was used to determine if the rating scale items on the *Teacher Quality Survey* clustered into meaningful

subscales that fall under the construct of teacher knowledge and skills for teaching academiclanguage to ELs (Turkan, 2012). This EFA was needed to determine the validity of the rating scale in addition to the number of outcome measures possible for conducting statistical analyses on the relationships among teacher quality inputs and processes (Goe, 2007) as they relate to ELs.

Assumptions to the EFA were met in terms of sample size, using KMO and Bartlett's test of Sphericity ($p < .001$) measures. All variables on the rating scale correlated at acceptable levels, and were normally distributed. A principal axis factor analysis was run using a varimax rotation although the rotation was not needed due to only one factor being retained. To illustrate, the factor analysis had eigenvalues over Kaiser's 1 and explained 73.042% of the variance. Inflexions in the scree plot justified retaining one factor, as the drop had virtually no steps in between the vertical line and plateau. Only a single factor was in the rating scale, which indicated that the construct of knowledge and skills in teaching academic language (Turkan et al., 2012) was captured using all 27 rating scale items. Toward those ends, a single variable, taken as the sum, was used in subsequent inferential statistical analyses as the dependent variable.

Implications and Recommendations – Rating Scale Items on the Teacher Quality Survey

Rating scale items used to measure how teachers report skill and knowledge on teaching EL students academic language were tested and found to be valid and reliable. While studies on teacher quality related to English learners used student outcomes to measure teacher effectiveness (Loeb et al., forthcoming; Masters et al., 2012), this study used items from an ETS study (Turkan et al., 2012) that assesses teacher quality in content instruction of ELs and adapted them to explore how teachers in Alabama report their current level of practice in this area. These

rating scale items served as the first tools for investigating the black box that records what happens in today's classrooms through the voices of general classroom teachers, intimating at how practices are used to educate ELs. For future studies, however, the rating scale items could be shortened as they load significantly on a single factor. The sections that follow deliberate on the findings from this study with regard to these reported measures of practice so that we can better understand teacher practice and consider logical next steps for future research and policymaking.

Teacher Quality Survey Results

A discussion of findings for research questions two through four will be presented:

2. To what extent do teachers report that they are prepared for teaching ELs academic language?
3. To what extent do teachers report they have adequate skills to teach ELs?
4. To what extent do (a) preference working with EL populations based on size, (b) in-service preparation hours in EL training, (c) teacher certification, and (d) the number of ELs taught in the classroom predict teacher reported skill level for teaching academic language?

Research Question Two: Teacher Inputs-Qualifications

Credentials and experience. Over half of teachers who participated in the survey have a master's degree in education, are certified in elementary education, and teach at the primary and intermediate levels (Pre-K–5). The rest of the participant sample was spread out in grades 6 through 12. The target sample was teachers working in Alabama schools with 20 or more ELs, and the majority of EL students are enrolled in the elementary level (Alabama State Department

of Education, personal communication, October 10, 2014), which might explain why there were fewer respondents who reported teaching at the secondary levels.

Descriptive data indicate that 58% of respondents have had little (0–10 hours) training related to English learners, whereas a small minority (8%) reported having between 31 or more hours of EL training. Given this study focused on teachers serving in schools with 20 or more ELs where the population may have greater implications for accountability, there is a need to increase the number of teachers who receive a substantial amount of hours in EL training. State level data indicate the majority of ELs in Alabama are concentrated in the elementary levels.

Implications. English learners spend the majority of their time in general education classrooms; however, the majority of teacher participants reported receiving only between zero to 10 hours of EL training. District and school leadership may not sense the urgency for EL training given the multitude of initiatives they are faced with prioritizing. Clearly, inclusive design of initiatives at the state level is imperative for streamlining how teachers are trained.

Research Question Three: Adequate Skills to Teach ELs

Areas where teachers reported feeling low to somewhat skilled were cognate skill building, critical thinking and cultural diversity, and knowledge of discourse of academic texts which uses passive voice to explain cause and effect. At the same time, teachers reported having higher knowledge and skills in providing oral and visual support for productive (written/spoken) language, teaching academic vocabulary in context, and strategic grouping configurations as their higher skill set area although the latter two were considerably less than half of teachers. Considering these results were gathered using descriptive data, research is needed to determine how accurately teachers have reported their skill set, which would require data triangulation through observation and archival data review. Specifically, the question that must be addressed

is the extent to which teachers are actually implementing these reported high and low skill sets so that the proper training supports, both at the in-service and teacher preparation levels, can be provided by district and state leadership support systems.

Toward that end, teachers in Alabama have had multiple opportunities provided at the local and state levels to participate in sustained training related to the areas where teachers reported their highest skill set. To illustrate, the three largest means of the rating scale items are associated with sheltered instruction strategies, skills that are known to effectively increase achievement levels through the development of academic language (August et al., 2009; Echevarria, et al., 2011). Sheltered instruction is an approach that makes academic content more accessible to English learners through a systematic process that takes a considerable amount of time to fully develop and implement (Echevarria, Vogt, & Short, 2012).

It is critical to establish whether teachers have been learning through these specific EL training avenues, or whether other state training initiatives have overlapped with sheltered instruction training, or possibly both? A positive problem to consider, because if teachers did not report higher knowledge and skill in this area on average, there is a strong possibility EL students would be that much more behind academically.

A more vexing issue concerns the linguistic domain where classroom teachers require intense, sustained support. As discussed in Chapter 3, the rating scale items were divided into two domains: pedagogical and linguistic. Toward that end, the bottom two of three means fell under the linguistic domain, and the third lowest under pedagogical. Although the latter rating scale item was categorized as pedagogical, it addressed culture as it plays into critical thinking. Culture is a cornerstone in applied linguistics and could also be classified within the linguistic domain, given that language and culture are inseparable constructs (Baker, 2006). The two other

linguistic items included cognate skill building and knowledge of discourse of academic texts that uses passive voice to explain cause and effect.

Implications. It appears that an area that warrants more exploration is the linguistic domain in terms of instructional skills for teaching ELs. While teaching sheltered instruction may have some overlap with other areas of training, such as reading comprehension and vocabulary development, a stronger focus on linguistic skills is needed to support the development of academic language for students learning English. With a national swing toward a common framework of instructional standards, i.e. Common Core State Standards, the expectation for all students lies in mastering complex text and discourse in math, English/language arts, and science. English learners now more than ever will need to be in classrooms where teachers have mastered the knowledge and skills needed to teach these more complex linguistic skills that prepare students for college and career.

Research Question Four

Significant indictors of teacher quality related to instructional practices for teaching English learners are teacher certification, teacher preference working with EL populations, hours of EL training, and prior experience teaching ELs in terms of how they are associated with teacher reported knowledge and skills for teaching EL students academic language. When analyzed in separate multiple linear regression models and in a combined hierarchical multiple regression, these indicators emerged as significant predictors of teacher rated skill level. From this study, we know that, in order of importance:

- Certification in ESL/ESOL and early childhood education matter more than certification in elementary education;

- Preference for teaching an EL population of small to none resulted in significantly lower skills rating than teachers who report preferring to teach in large EL contexts;
- Teachers who reported having zero to ten hours of EL professional development tend to report a considerably lower practice and skill level than those who report having at least 40 or more hours of professional development related to teaching English learners;
- Teachers who reported having larger numbers of EL students during the prior school year had a significant, positive effect on teacher reported skill level although causation was not determined.

Certification. Compared to certification in elementary education, certification in either early childhood education, or ESL/ESOL in addition to a core subject area (or elementary) matters in terms of how high a teacher reports his or her skill level to teach EL academic language. Teachers certified in elementary education were predicted to score lower on skill level than teachers certified in early childhood or ESL/ESOL in combination with another area of certification. Thus certification is a significant indicator of teacher quality for EL students. Surprisingly, the only positive correlation among content area certification and skill level found was certification in English/Language Arts, which may account for content teachers not assuming their roles as language educators.

Implications. In Alabama, the majority of EL students are enrolled in the primary and intermediate levels; however, in this study teachers who reported having additional certification in ESL/ESOL tended to report a considerably higher skills rating than teachers certified in elementary education alone, meaning there is a population of teachers in the state who can potentially deliver adequate services to their EL students, whereas others reportedly lack the

skills and resources to be able to do so. These findings speak to current state policy, as teachers in Alabama who have elementary certification are permitted to serve as the ESL teacher—the language teacher. This means that it is permissible for districts and schools in Alabama to place EL students with regular elementary classroom teachers for ESL services instead of with an ESL certified teacher, as they are in short supply. Yet elementary certified teachers reported practicing and knowing considerably less about these knowledge and skills than teachers who have additional certification in ESL/ESOL. Findings indicate that certification in elementary education alone is inadequate in terms of the level of teacher quality needed to educate English learners. There may be promise if these teachers are certified in early childhood education if no ESL certified teachers are accessible. And if this is the case, the compelling question is, how are early childhood elementary teacher certification programs preparing teachers? Further research is needed in this area of EL students and teacher certification.

Preference Working with EL Population by Size

Other significant predictors of teacher reported skill level include preference working with EL population size. In their study, Loeb et al. (forthcoming) found that teacher preference for working with the EL subgroup predicted teacher effectiveness in math. Those findings were corroborated in this study insofar as they predicted teacher reported knowledge and skill level. Compared to teachers who prefer working with large EL populations, teachers who preferred working with either small to no EL populations were predicted to report significantly lower skills on the rating scale, and shared a negative relationship with reported skill level.

Implications. Administrators who are responsible for hiring teachers must ask the right questions during the interview process, and they must ensure that EL students are placed with teachers who are confident in their ability to work with English learners. Reporting a preference

for working with a small to zero population of ELs does not necessarily mean teachers do not like these students; rather, they may not have the adequate skills and training needed to work with them. As such, they may tend to rate themselves lower in skill than those with more intensive training and preparation. While there may be a small minority who truly does not want to teach linguistically and culturally diverse students, that cannot not be the default interpretation of these findings. Clearly, lack of training can have a powerful impact on teacher preference, and the implications are for instructional leaders to plan and act accordingly.

Hours of EL Professional Development

Hours of EL professional development clearly matters in terms of teacher quality for ELs. In this study, teachers were able to identify with the knowledge and skills needed to develop academic language for ELs on the rating scale and evaluate their practice at higher levels than if they had not received a significant amount of EL training. The more training hours related to teaching EL students reported by teachers, the higher their reported skill level in teaching ELs academic language, yet a majority of respondents reported having less than 10 hours of EL training. Findings suggest that teachers with 40 hours or more of EL training tend to report a significantly higher level of knowledge and skill level.

Implications. It is likely that most teachers working in elementary schools with 20 or more EL students in Alabama have not had a sufficient number of training hours associated with the higher knowledge and skill base EL students need to be successful academically. This deficit in training means the practices required for teaching these students academic language are not well known and/or implemented in the classroom with a sufficient level of skill. Too often, instructional leadership has provided general classroom teachers with a one-shot dosage of EL professional development, and this approach clearly is not working with regard to what teachers

report knowing and doing in the classroom to support EL students learning academic language. Without having this knowledge base from which to put into practice (the processes function of the teacher quality framework as per Goe, 2007) students do not have access to the core curriculum, and consequently, the achievement gap does not close.

Moreover, the question arises as to whether the majority of teacher evaluators, most likely school administrators, also lack this level of knowledge needed to evaluate teachers with reliability and validity. Designs in widely used observation protocols and evaluation rubrics lack these very specific constructs of linguistic instructional practices that impact the EL subgroup (Jones et al., 2013; Kane et al., 2012; Milanowski, 2004; Turkan et al., 2012), so how and to what extent do teacher evaluators acquire these knowledge and skills? The problem of teachers not receiving the appropriate level of training is compounded by teacher evaluators who are also lack training. These issues surrounding teacher quality have major implications for policy and decision-making concerning teacher evaluation processes, both in terms of student growth and observation measures.

Experience with EL Population Per Year

Most teacher respondents reported having approximately seven EL students during the previous school year, which matters in terms of teacher quality related to English learners. A significant, positive relationship was found between the reported number of EL students taught this school year and last school year with their reported skill level. Findings from this study may support previous studies (Master et al., 2012) that suggest teachers with prior experience teaching at least six ELs in the classroom is highly predictive of math EL achievement. However, when all indicators (e.g., certification, hours of EL PD, preference) were added together in the HMR analysis, experience with EL population entered last, neither predictor

(number this year and number last year) made a significant contribution to the model although correlations for both variables and skill level were moderately positive at the significant levels for both the MLR and HMR analyses.

Implications. The more practice teachers have working with ELs in the classroom the higher they tend to rate their skill set for teaching ELs. This finding may be the result of having attended more intensive trainings related to this student population in conjunction with more experience working with them. Implications for instructional leaders relate directly to the hiring process and in the placement of EL students when they enroll in school. Leaders in schools with significant EL populations need to know whether their potential teaching candidates have had experience working with larger numbers of EL students, as this indicator may act as a lever for closing the achievement gap. Additionally, when these students arrive for enrollment, a well-experienced faculty in terms of prior experience with ELs should sustain the growing numbers of the EL population.

Recommendations for Policy, Leadership, and Practice

This study is relevant to education because examining the teacher knowledge and skills gap related to teaching ELs academic language establishes the degree to which teachers currently report addressing the kinds of instructional practices these students need to achieve in English speaking classrooms (Baker et al., 2014; Turkan et al., 2012). The role of teacher processes in the teacher quality framework (Goe, 2007) cannot be left out of the equation for solving this problem. The expectation for classroom teachers to teach EL students academic language is misguided when evaluation frameworks and teacher standards provide them with inadequate supports and resources for getting the job done. Findings from this study allow researchers,

policymakers, and practitioners to move forward with new information that spotlights the practices teachers in Alabama need to be quality teachers of EL students.

The following indicators mattered most in this study when assessing teacher quality for teaching English learners:

- Certification in either early childhood education, or the combination of ESL/ESOL and another certification area (such as elementary education);
- Preference working with mid-to-large EL populations versus none to small, or no preference;
- 40 hours or more of professional development related to the instruction of ELs versus zero to twenty;
- Prior experience teaching five or more EL students at a time in the classroom.

These indicators are not an exhaustive list, as much more research is needed in the area of teacher quality studies related to English learners. However, although certification and having more than 40 hours of EL training are the most promising indicators, a combination of these factors may support this population of students in the elementary setting. If leaders are hard pressed to find teachers who fit all of these descriptions, they must take steps to provide them with supports and resources to meet these criteria through sustained professional development, which can include in-service hours or coursework, ideally leading to certification or an advanced degree in ESL/ESOL.

Policymakers and stakeholders are presented with the opportunity to reform teacher evaluation systems as they redesign current appraisal systems. During this window of opportunity, both groups should evaluate their educational context and needs in terms of teacher preparation requirements and professional development models and make needed adjustments to

how teachers are evaluated. First, I recommend policymakers require additional teacher certification in ESL/ESOL where there are significant populations of English learners present, given its bearing on teacher reported knowledge and skills. Elementary certification alone is inadequate for serving this population, yet the majority of EL students are enrolled in this level. Further, given the possibility that certification in early childhood may have a positive relationship with teacher reported skill, institutes of higher education should take steps to expand the curriculum content to include more knowledge and skills in teaching academic language to culturally and linguistically diverse learners. Offering a particular strand or course, much like the program Auburn University implemented in 2014, that emphasizes linguistic diversity may show promise for this population of students, not to mention prospects for teacher candidates and a larger pool of qualified applicants.

Next, for in-service teachers who opt not to return to school for additional certification, instructional leaders must provide them with a sustained model of professional development where they can earn more than 40 hours of EL training. While this model conforms to current federal mandates for districts receiving Title III funds, this study found that the majority of teachers working in schools with 20 or more EL students have received less than 10 hours and were unable to report adequate levels of knowledge and skills in teaching ELs academic language, the language needed to close the achievement gap. Creating awareness and eliciting care among leadership is imperative, as laws and regulations are not powerful enough to create change in practice. Leaders need to know what works for all students they serve and provide the support and tools teachers need to be effective practitioners for academic language learners.

Toward those ends, awareness training for principals that emphasizes the need for sustained teacher training and hiring faculty with substantive experience working with EL

students is needed. Teachers should be placed strategically in settings where their skills and talent are needed. Further, educational leadership programs should include a curriculum component that prepares leaders to understand and address the issues surrounding English learner students.

Recommendations for Future Research

Examining teachers' voices on their current level of knowledge and skill in teaching EL students academic language provides us with more insight into teacher quality for this subgroup. Most teachers who participated in this study have not been trained nor are prepared to teach EL students and reported having lower knowledge and skills than their counterparts. A logical next step is to observe classrooms where teachers have reported having substantive skills in teaching academic language and observe how they and their students perform with regard to these skill sets. Next, exploring how critical aspects of skills for teaching academic language to ELs with teacher evaluation tools currently being piloted at state and local levels, especially in areas where there are substantial populations of EL students, function as tools for professional growth.

Another needed area of research should explore how principals currently evaluate general classroom teachers who serve EL students. For example, to what extent are principals aware of the evidence-based knowledge and skills teachers require for teaching academic language to ELs, and how do they report their level of preparation to lead in schools with culturally and linguistically diverse student populations? How much training and support do they currently provide their faculty, and how much do they seek out for themselves?

Concluding Remarks

In terms of policy, teacher quality and evaluation have been defined by and currently favor using student outcomes as quantitative measures of teacher effectiveness. While there is a

place for these measures—such as using them in combination with multiple measures—research tells us to proceed with caution. For classroom teachers who serve English learners, more factors unique to this special population exist and have implications that render these metrics questionable in how they impact teacher evaluation. Policy regarding teacher quality and evaluation crafted with deliberate thinking and careful attention to evidenced-based teacher practices versus student outcomes may prove favorable with regard to how we evaluate teachers who serve ELs and ultimately with closing the achievement gap between them and their English speaking counterparts. What teachers know and do in the classroom to support academic achievement of English learners are the practices that warrant more examination and more and emphasis in how we approach teacher quality and evaluation.

REFERENCES

Aarnoutse, C. A., Van den Bos, K. P., & Brand-Gruwel, S. (1998). Effects of listening comprehension training on listening and reading. *The Journal of Special Education, 32*(2), 115–126.

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

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1–28.

Alabama State Department of Education. (2014). English learners policy and procedures manual. Retrieved from http://alex.state.al.us/ell/?q=node/27

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of the literature on academic English: Implications for K–12 English language learners*. Arlington, VA: The George Washington University Center for Equity and Excellence in Education.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness, 2*(4), 345–376. doi:10.1080/19345740903217623

August, D. E., & Shanahan, T. E. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Lawrence Erlbaum Associates Publishers.

Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Tonawonda, NY: Multilingual Matters.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. EPI Briefing Paper# 278. Economic Policy Institute.

Baker, S. K., Gersten, R., Haager, D., & Dingle, M. (2006). Teaching practice and the reading growth of first-grade English learners: Validation of an observation instrument. *The Elementary School Journal, 107*(2), 199–220.

Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., ... & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school*. IES Practice Guide. NCEE 2014-4012. What Works Clearinghouse.

Ballenger, B. (1997). Social identities, moral narratives, scientific argumentation: Science talk in a bilingual classroom. *Language & Education, 11*(1), 1–14. doi:10.1080/09500789708666715

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy, 54*(1), 4–12.

Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. *Handbook of Research on Teaching the English Language Arts, 2*, 752–785.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251–271.

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). *The effects and uses of diverse vocabulary instructional techniques*.

Betts, J. R., Zau, A., & Rice, L. (2003). *Determinants of student achievement: New evidence from San Diego* (pp. 1–5821). San Francisco: Public Policy Institute of California.

Blanton, L. P., Sindelar, P. T., & Correa, V. I. (2006). Models and measures of beginning teacher quality. *The Journal of Special Education, 40*(2), 115–127.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106*(1), 3–20.

Bradley, M. C., Daley, T., Levin, M., O'Reilly, R., Parsad, A., Robertson, A., & Werner, A. (2011). *IDEA national assessment implementation study* (NCEE 2011-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education. Retrieved from http://ies.ed.gov/ncee/ pubs/20114026/pdf/20114027.pdf

Braun, H. I. (2005). Using student progress to evaluate teachers: A primer on value-added models. Policy Information Center, Mail Stop 19-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001. Available: www.ets.org/research/pic

Brown, B., Ryoo, K., & Rodriguez, J. (2010). Pathway towards fluency: Using "disaggregate instruction" to promote science literacy. *International Journal of Science Education, 32*(11), 1465–1493. doi:10.1080/09500690903117921

Burns, D. (2011). *Examining the effect of an overt transition intervention on the reading development of at-risk English-language learners in first grade* (Unpublished doctoral dissertation). University of Oregon, Oregon.

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability measurement challenges for students with disabilities and recommendations for research. *Educational Researcher, 39*(7), 537–544.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103–127.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D., . . . White, C. E. (2004). Closing the gap: Addressing the vocabulary needs for English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188–215. doi:10.1598/RRQ.39.2.3

Cavalluzzo, L. C. (2004). *Is national board certification an effective signal of teacher quality*? CNA Corporation.

Chamot, A. U. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching, 1*(1), 14–26.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*(6), 673–682.

Coady, M. R., de Jong, E. J., & Harper, C. (2011). Quality teacher preparation for ELLs: Preliminary findings from Florida. Professional development in action: Improving teaching for English learners, 11(4).

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21*(4), 617–641.

Collier, V. P., & Thomas, W. P. (1989). How quickly can immigrants become proficient in school English. *Journal of Educational Issues of Language Minority Students, 5*(Fall), 26–39.

Culbertson, M. *What's the Big Deal About Missing Data?* (PDF document). Retrieved from http://teams.mspnet.org/media/data/TEAMS_APRIL_WEBINAR.pdf?media_000000008320.pdf

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In *Encyclopedia of language and education* (pp. 487–499). US: Springer.

Danielson, C. (2013). The framework for teaching. Evaluation Instrument. The Danielson Group

De Jong, E. J., & Harper, C. A. (2005). Preparing mainstream teachers for English-language learners: Is being a good teacher good enough? *Teacher Education Quarterly*, 101–124.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 158–165.

Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research & Practice*, *23*(2), 79–89. doi:10.1111/j.1540-5826.2008.00266.x

DiCerbo, P. A., Anstrom, K. A., Baker, L. L., & Rivera, C. (2014). A review of the literature on teaching academic English to English language learners. *Review of Educational Research*, 0034654314532695.

Duke, N. K., Pressley, M., Hilden, K., Stone, C. A., & Silliman, E. R. (2004). Difficulties with reading comprehension. *Handbook of Language and Literacy: Development and Disorders*, 501–520.

Echevarria, J., Richards-Tutor, C., Chinn, V. P., & Ratleff, P. A. (2011). Did they get it? The role of fidelity in teaching English learners. *Journal of Adolescent & Adult Literacy*, *54*(6), 425–434.

Echevarria, J., Short, D., & Powers, K. (2006). School reform and standards-based education: A model for English-language learners. *The Journal of Educational Research*, *99*(4), 195–211.

Echevarria, J., Vogt, M. J., & Short, D. J. (2012). *Making content comprehensible for English learners: The SIOP model* (4th edition). New York: Pearson Education.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). *Do teachers' race, gender, and ethnicity matter?: Evidence from NELS88* (No. w4669). National Bureau of Economic Research.

Fields, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks: Sage.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*(1), 174–206. doi:10.3102/00028312034001174

Gallagher, H. A. (2002). *The relationship between measures of teacher quality and student achievement: The case of Vaughn Elementary* (Unpublished doctoral dissertation). University of Wisconsin-Madison.

Gersten, R. (1996). The double demands of teaching English language learners. *Educational Leadership, 53*(5), 18–22.

Gersten, R., & Geva, E. (2003). Teaching reading to early language learners. *Educational Leadership, 60*(7), 44–49.

Gersten, R., Baker, S. K., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners. *Remedial & Special Education, 26*(4), 197–206.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. National Comprehensive Center for Teacher Quality.

Goe, L., & Stickler, L. M. (2008). *Teacher quality and student achievement: making the most of recent research*. TQ Research & Policy Brief. National Comprehensive Center for Teacher Quality.

Goldhaber, D. D., & Brewer, D. J. (1996). *Evaluating the effect of teacher degree level on educational performance*.

Gomez, R., Parker, R., Lara-Alecio, R., & Gomez, L. (1996). Process versus product writing with limited English proficient students. *The Bilingual Research Journal, 20*(2), 209–233. doi:10.1080/15235882.1996.10668628

Good, R. H., Kaminski, R. A., Smith, S., Simmons, D. S., Kame'enui, E. J., & Wallin, J. (2003). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. *Reading in the classroom: Systems for observing teaching and learning*. Baltimore: Paul H. Brookes.

Gottardo, A. (2002). The relationship between language and reading skills in bilingual Spanish-English speakers. *Topics in Language Disorders, 22*(5), 46–70.

Gottlieb, M., & Ernst-Slavit, G. (2014). *Academic language in diverse classrooms: definitions and contexts*. Corwin Press.

Graves, M. F. (2000). A vocabulary program to complement and bolster a middle-grade comprehension program. *Reading for meaning: Fostering comprehension in the middle grades*, 116–135.

Griffin, P., Burns, M. S., & Snow, C. E. (Eds.). (1998). Preventing reading difficulties in young children. National Academies Press.

Hakuta, K., & August, D. (Eds.). (1998). *Educating language-minority children*. National Academies Press.

Halliday, M. A. K., & Hasan, R. (1989) *Language, context, and text: Aspects of a language in a social-semiotic perspective* (2nd ed.). Oxford: Oxford University Press.

Hampton, E., & Rodriguez, R. (2001). Inquiry science in bilingual classrooms. *Bilingual Research Journal, 25*(4), 461–179.

Hanushek, E. A., & Rivkin, S. G. (2010). The quality and distribution of teachers under the No Child Left Behind Act. *The Journal of Economic Perspectives*, 133–150.

Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics., 4*(1), 131–157.

Heck, R. H. (2007). Examining the relationship between teacher quality as an organizational property of schools and students' achievement and growth rates. *Educational Administration Quarterly, 43*(4), 399–432.

Heneman III, H. G., Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge-and skill-based pay*. CPRE Policy Briefs. RB-45. Consortium for Policy Research in Education.

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. Research & Policy Brief. National Comprehensive Center for Teacher Quality.

Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2014). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism*, 1367006914521698.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101–136.

Johnson, E., & Semmelroth, C. L. (2014). Special education teacher evaluation why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Intervention, 39*(2), 71–82.

Jones, N. D., & Brownell, M. T. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention, 39*(2), 112–124.

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher, 42*(4), 234–241. doi:10.3102/0013189X12468211

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Kearsey, J., & Turner, S. (1999). The value of bilingualism in pupils' understanding of scientific language. *International Journal of Science Education, 21*(10), 1037–1050. doi:10.1080/095006999290174

Kennedy, M. M. (1999). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis, 21*, 345–363.

Kieffer, M. J., & Lesaux, N. K. (2012). Effects of academic language instruction on relational and syntactic aspects of morphological awareness for sixth graders from linguistically diverse backgrounds. *The Elementary School Journal, 112*(3), 519–545. doi:10.1086/663299

Kim, J., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., . . . Land, R. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed Latino English language learners in grades 6 to 12. *Journal of Research on Educational Effectiveness, 4*(3), 231–263. doi:10.1080/19345747.20 10.523513

Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice, 26*(3), 11–20.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom*. Michigan: Pergamon Press.

Lakin, J. M., & Young, J. W. (2013). Evaluating growth for ELL students: implications for accountability policies. *Educational Measurement: Issues and Practice, 32*(3), 11–26.

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39*(6), 1005.

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196–228. doi:10.1598/ RRQ.45.2.3

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents' evidence from a randomized field trial. *American Educational Research Journal*, 0002831214532165.

Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. T. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities, 39*(5), 390–398.

Loeb, S., & Candelaria, C. (2012). *How stable are value-added estimates across years, subjects, and student groups?* Carnegie Knowledge Network.

Loeb, S., Soland, J., & Fox, L. (forthcoming). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Education Evaluation and Policy Analysis*.

Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education preparing classroom teachers to teach English language learners. *Journal of Teacher Education, 59*(4), 361–373.

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2012). *Different skills? Identifying differentially effective teachers of English language learners*. Manuscript submitted for publication.

Mathes, P. G., Pollard-Durodola, S. D., Cárdenas-Hagan, E., Linan-Thompson, S., & Vaughn, S. (2007). Teaching struggling readers who are native Spanish speakers: What do we know? *Language, Speech, and Hearing Services in Schools, 38*(3), 260–271.

Mendro, R. L. (1998). Student achievement and school and teacher accountability. *Journal of Personnel Evaluation in Education, 12*(3), 257–267.

Meyer, J. P., Cash, A. H., & Mashburn, A. (2011). Occasions and the reliability of classroom observations: Alternative conceptualizations and methods of analysis. *Educational Assessment, 16*(4), 227–243.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33–53.

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.

Munoz, M. A., Prather, J. R., & Stronge, J. H. (2011). Exploring teacher effectiveness using hierarchical linear models: Student-and classroom-level predictors and cross-year stability in elementary school reading. *Planning and Changing*, *42*, 241–273.

Muñoz, M. A., Scoskie, J. R., & French, D. L. (2013). Investigating the "black box" of effective teaching: the relationship between teachers' perception and student achievement in a large urban district. *Educational Assessment, Evaluation and Accountability, 25*(3), 205–230.

Nagy, W. E., McClure, E. F., & Mir, M. (1997). Linguistic transfer and the use of context by Spanish-English bilinguals. *Applied Psycholinguistics, 18*(04), 431–452.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*(1), 91–108.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. TESOL Journal, 9(2), 6–10.

National Clearinghouse for English Language Acquisition. (2011). *The growing number of English learner students 2009–10*. Retrieved September 6, 2014 from http://www.ncela.us/files/uploads/9/growing_EL_0910.pdf

National Comprehensive Center for Teacher Quality. (2012). *Summary of expert forum on the evaluation of teachers of English language leaners*. Retrieved at: http://www.gtlcenter.org/sites/default/files/docs/ForumSummary_July2012.pdf

National Reading Panel (U.S.), National Institute of Child Health, & Human Development (U.S.). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its*

implications for reading instruction: Reports of the subgroups. National Institute of Child Health and Human Development, National Institutes of Health.

Nelson, J., Vadasy, P., & Sanders, E. (2011). Efficacy of a tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research, 43*(2), 184–211. doi:10.1177/1086296X11403088

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Peercy, M. M. (2011). Preparing English language learners for the mainstream: Academic language and literacy practices in two junior high school ESL classrooms. *Reading & Writing Quarterly, 27*(4), 324–362.

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.

Rader-Brown, L., & Howley, A. (2014). Predictors of the instructional strategies that elementary school teachers use with English language learners. *Teachers College Record*, *116*, 050306.

Ransford-Kaldon, C. R., Flynt, E. S., Ross, C. L., Franceschini, L., Zoblotsky, T., Huang, Y., & Gallagher, B. (2010). *Implementation of effective intervention: An empirical study to evaluate the efficacy of Fountas & Pinnell's Level Literacy Intervention System* (LLI) 2009–2010. Memphis, TN: University of Memphis, Center for Research in Educational Policy. Retrieved from http://www.mem- phis.edu/crep/pdfs/lli-efficacy.pdf

Reeves, J. R. (2006). Secondary teacher attitudes toward including English-language learners in mainstream classrooms. *The Journal of Educational Research, 99*(3), 131–143.

Rivera, M. O., Francis, D. J., Fernandez, M., Moughamian, A. C., Jergensen, J., & Lesaux, N. K. (2010). *Effective practices for English language learners: principals from five states speak*. Center on Instruction.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *The Teachers College Record*, *104*(8), 1525–1567.

Ross, M. E. & Shannon, D. M. (2011). *Handbook of applied quantitative methods in education*. Dubuque, IA: Kendall Hunt.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, 103–116.

Ryoo, K. (2009). *Learning science, talking science: The impact of a technology-enhanced curriculum on students' science learning in linguistically diverse mainstream classrooms* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3364450)

Salahshour, F., Sharifi, M., & Salahshour, N. (2013). The relationship between language learning strategy use, language proficiency level and learner gender. *Procedia-Social and Behavioral Sciences, 70*, 634–643.

Samson, J. F., & Collins, B. A. (2012*). Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness*. Center for American Progress.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57–67.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). Comparison of the *effects of NBPTS certified teachers with other teachers on the rate of student academic progress*. Final Report. National Board for Professional Teaching Standards.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Research Report. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Routledge.

Short, D. J., Echevarría, J., & Richards-Tutor, C. (2011). Research on academic literacy development in sheltered instruction classrooms. *Language Teaching Research, 15*(3), 363–380.

Short, D. J., Fidelman, C. G., & Louguit, M. (2012). Developing academic language in English language learners through sheltered instruction. TESOL Quarterly, 46(2), 334–361.

Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners – A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

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

Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101*(2), 305–314. doi:10.1037/ a0014217

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. *The Cambridge Handbook of Literacy*, 112–133.

Solari, E. J., & Gerber, M. M. (2008). Early comprehension instruction for Spanish-speaking English language learners: Teaching text-level reading skills while maintaining effects on word-level skills. *Learning Disabilities Research & Practice, 23*(4), 155–168. doi:10.1111/j.1540-5826.2008.00273.x

Strong, M., Gargani, J., & Hacifazlioğlu, Ö. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education, 62*(4), 367-382.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. Journal of Teacher Education, 62(4), 339–355.

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, *20*(3–4), 165–184.

Téllez, K., & Waxman, H. C. (2005). *Quality teachers for English language learners*. *A review of the research*. Laboratory for Student Success (LSS), The Mid-Atlantic Regional Educational Laboratory.

Turkan, S., & Buzick, H. M. (2014). Complexities and issues to consider in the evaluation of content teachers of English language learners. *Urban Education*, 0042085914543111.

Turkan, S., Croft, A., Bicknell, J., & Barnes, A. (2012). *Assessing quality in the teaching of content to English language learners*. Research Report, ETS RR-12-10. Princeton: New Jersey.

Turkan, S., De Oliveira, L. C., Lee, O., & Phelps, G. (2014). Proposing a knowledge base for teaching academic content to English language learners: Disciplinary linguistic knowledge. *Teachers College Record, 116*(030308), 1–30.

U.S. Department of Education. (2001). *No Child Left Behind*. Washington, DC: Author. Retrieved June 6, 2014, from http://www2.ed.gov/policy/elsec/leg/esea02/index.html

U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Washington, DC: Author. Retrieved April 23, 2014, from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2010). *School Improvement Grants*. Washington, DC: Author. Retrieved April 6, 2014, from http://www2.ed.gov/programs/sif/index.html

U.S. Department of Education. (2012). ESEA Flexibility. Washington, DC: Author. Retrieved April 25, 2014 from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sanchez, E. (2014). Core Academic Language Skills (CALS): An expanded operational construct and a novel instrument to chart school-relevant language proficiency in pre-adolescent and adolescent learners. *Applied Psycholinguistics xx*, no. x:xx-xx. Retrieved September 7, 2014 at http://nrs.harvard.edu/urn-3:HUL.InstRepos:11380186

Vaughn, S., Cirino, P. T., Linan-Thomp- son, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., . . . Francis, D. J. (2006). 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. doi:10.3102/00028312043003449

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional children*, *69*(4), 391–409.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New Teacher Project.

Appendix 1

Teacher Survey

Directions: Thank you for taking the time to complete this survey. The purpose of this research study is to gather information about general classroom teacher practices for English learners (ELs), students who are limited in English proficiency (LEP) and placed in an English language instruction education program.

When responding to the survey, please consider only LEP 1 and LEP 2 students. LEP 1 students are in their first year in a US school, and LEP 2 students are in their second year or more in a US school. The purpose of this study does not concern Former LEP (FLEP) students who have exited the English language instruction education program are on monitoring status.

This survey is relevant to all general classroom teachers who currently teach in the classroom and work in schools with 20 or more English learners, even if they do not currently have ELs in their classroom. This survey is not intended for ESL/ESOL teachers.

This survey will take approximately 10 minutes to complete.

PART 1 In what area/s are you certified to teach? Check all that apply.

- ❑ Elementary Education
- ❑ Reading
- ❑ Early Childhood Education
- ❑ Mathematics
- ❑ Science
- ❑ Social Studies
- ❑ English/Language Arts
- ❑ English as a Second Language (ESL/ESOL)
- ❑ Other (please specify in text box) _____

2 What is your current primary teaching assignment? Check all that apply.

- ❑ Elementary Education
- ❑ Reading
- ❑ Mathematics
- ❑ Science
- ❑ Social Studies
- ❑ English Language/Arts
- ❑ Elective/Special (please specify in text box below) _____
- ❑ Other (please specify in text box below) _____

3 What grade level do you currently teach? Check all that apply.

- ❑ Pre-K
- ❑ Kindergarten
- ❑ 1st
- ❑ 2nd
- ❑ 3rd
- ❑ 4th
- ❑ 5th
- ❑ 6th
- ❑ 7th
- ❑ 8th
- ❑ 9th
- ❑ 10th
- ❑ 11th
- ❑ 12th

4 Enter the number of years of teaching experience you have in the box below.

5 Select each type of advanced degree you have earned.

- ❑ Master's degree in ESL/TESOL
- ❑ Master's degree in Education
- ❑ Education Specialist (EdS). List area in text box below. _____
- ❑ Doctorate (PhD or EdD). List area in text box below. _____
- ❑ Not applicable
- ❑ Other (please describe in text box below) _____

Answer: Select each type of degree you have earned. Master's degree in ESL/TESOL Is Not Selected

6 Have you had any university coursework in ESL/ESOL?

- ❍ Yes
- ❍ No

7 How many hours of professional development training, including continuing education units (CEUs), have you earned in the last 5 years where the primary topic was teaching English learners?

- ❍ Zero hours
- ❍ 1 - 10 hours
- ❍ 11 - 20 hours
- ❍ 21 - 30 hours
- ❍ 31 - 40 hours
- ❍ More than 40 hours

8 Select the largest number of English learners that you have had in a class at any point last school year.

- ❍ 0
- ❍ 1
- ❍ 2
- ❍ 3
- ❍ 4
- ❍ 5
- ❍ 6
- ❍ 7
- ❍ 8
- ❍ 9
- ❍ 10
- ❍ More than 10
- ❍ Not applicable

9 Select the largest number of English learners that you have had in a class at any point during the current school year.

- ❍ 0
- ❍ 1
- ❍ 2
- ❍ 3
- ❍ 4
- ❍ 5
- ❍ 6
- ❍ 7
- ❍ 8
- ❍ 9
- ❍ 10
- ❍ More than 10

10 PART 2 How skillful are you at using the pedagogical skills and practices that are listed in the left-hand column for teaching English learners academic language?

		Less Skilled	Somewhat Skilled	Skilled	Very Skilled	Highly Skilled
Make abstract concepts accessible to ELs.	❍	❍	❍	❍	❍	❍
Draw upon ELs' cultural and educational background to facilitate learners' comprehension and discussion of academic texts.	❍	❍	❍	❍	❍	❍
Teach metacognitive language-learning strategies (e.g., steps in problem solving, acquiring study skills) to ELs.	❍	❍	❍	❍	❍	❍
Provide multiple opportunities for ELs to process content in group contexts, including with their English-speaking monolingual or proficient bilingual peers (e.g., peer work assignments, whole-class discussions).	❍	❍	❍	❍	❍	❍
Include language objectives alongside content objectives in planning lessons.	❍	❍	❍	❍	❍	❍

1 PART 2 (CONT.) How skillful are you at using the pedagogical skills and practices that are listed in the left-hand column for teaching English learners academic language?

	Less Skilled	Somewhat Skilled	Skilled	Very Skilled	Highly Skilled
Provide ELs with oral and visual support (e.g., think-alouds, word picture cards) to produce language orally and in writing.	❍	❍	❍	❍	❍
Provide alternative forms of assessments for ELs to gauge their comprehension and production of text.	❍	❍	❍	❍	❍
Implement various strategies to differentiate instruction for ELs' success in language and literacy development.	❍	❍	❍	❍	❍
Supplement curriculum and textbook materials with other sources to aid EL learning in content areas.	❍	❍	❍	❍	❍
Teach ELs new vocabulary in context, both intentionally and incidentally.	❍	❍	❍	❍	❍
Apply various methods to incorporate different interactional and task engagement styles that ELs might bring from their cultural backgrounds into classroom work.	❍	❍	❍	❍	❍

12 PART 2 (CONT.) How skillful are you at using the pedagogical skills and practices that are listed in the left-hand column for teaching English learners academic language?

	Less Skilled	Somewhat Skilled	Skilled	Very Skilled	Highly Skilled
Develop higher order thinking skills for ELs at any English proficiency level through questioning and elicitation techniques.	❍	❍	❍	❍	❍
Encourage critical thinking by acknowledging ELs' diverse cultural experiences (e.g., arguments from authority may be highly valued in some cultures).	❍	❍	❍	❍	❍
Help ELs to understand discipline-specific concepts within a content area through the use of graphic organizers (e.g., semantic maps).	❍	❍	❍	❍	❍
Adapt texts (e.g., paraphrasing unfamiliar expressions) to make content-specific concepts accessible to ELs.	❍	❍	❍	❍	❍
Explicitly explain to ELs genres (e.g., expository, narrative) that are applicable to a specific content area.	❍	❍	❍	❍	❍

13 PART 2 (CONT.) How skillful are you at using the linguistic skills and practices that are listed in the left-hand column for teaching English learners academic language?

	Less Skilled	Somewhat Skilled	Skilled	Very Skilled	Highly Skilled
Build on ELs' knowledge of cognates between English and their home language.	❍	❍	❍	❍	❍
Distinguish between core content vocabulary and common everyday vocabulary.	❍	❍	❍	❍	❍
Pace oral communications and instructions appropriately for ELs' comprehension (e.g., repeat or clarify ideas, avoid or teach idioms and slang).	❍	❍	❍	❍	❍
Develop ELs' metalinguistic awareness of the English language (i.e., ability to distinguish between literal and implied meanings).	❍	❍	❍	❍	❍
Design written and oral activities to provide ELs at any English proficiency level with the opportunities to express their ideas and perspectives.	❍	❍	❍	❍	❍
Help ELs identify areas of difficulty in academic vocabulary during in-class discussions of textbook passages.	❍	❍	❍	❍	❍

14 PART 2 (Cont.) How skillful are you at using the linguistic skills and practices that are listed in the left-hand column for teaching English learners academic language?

	Less Skilled	Somewhat Skilled	Skilled	Very Skilled	Highly Skilled
Help ELs decode meaning from highly abstract and culturally embedded phrases and sentences by encouraging them to infer meaning from context.	❍	❍	❍	❍	❍
Transform textbook content into meaningful chunks for ELs (e.g., highlight key concepts).	❍	❍	❍	❍	❍
Scaffold ELs' ability to rephrase or paraphrase academic language in their own words.	❍	❍	❍	❍	❍
Know that the discourse of academic texts in content areas includes the use of passive voice in describing events or explaining cause and effect.	❍	❍	❍	❍	❍
Making explicit the transition from everyday oral language to more technical language or register of the content area.	❍	❍	❍	❍	❍

15 Part 3

	serves a large population of ELs.	serves a small population of ELs.	has no EL population.	no preference
Given a choice, I would choose to teach in a school that...	❍	❍	❍	❍

16 Are you fluent in the same language as any of your past or current English learners?
- ❍ Yes (list language in text box) _____
- ❍ No

17 Enter your age in the box below.

18 What is your gender?
- ❍ Male
- ❍ Female

19 Select your race from the options below.
- ❍ White/Caucasian
- ❍ African American
- ❍ Hispanic
- ❍ Asian
- ❍ Native American
- ❍ Pacific Islander
- ❍ Multi-racial

20 If you would like results to this study once it has been completed, please use this link to a new survey where you will be able to provide your email address. This way, your anonymity will be protected with the information you provide here. You may need to cut and paste the link to your web browser: https://auburn.qualtrics.com/SE/?SID=SV_9HzIDuCib9fnixD

Appendix 2

Descriptive Statistics for the Rating Scale Items

Rating Scale Item	Frequencies					
	M SD	*Less Skilled* No. (%)	*Somewhat Skilled* No. (%)	*Skilled* No. (%)	*Very Skilled* No. (%)	*Highly Skilled* No. (%)
1. Make abstract concepts accessible to ELs.	2.59 .949	22 (9.9)	75 (33.6)	71 (31.8)	25 (11.2)	6 (2.7)
2. Draw upon ELs' cultural and educational background to facilitate learners' comprehension and discussion of academic texts.	2.48 .985	28 (12.6)	83 (37.2)	64 (28.7)	17 (7.6)	9 (4.0)
3. Teach metacognitive language-learning strategies (e.g., steps in problem solving, acquiring study skills) to ELs.	2.54 .964	25 (11.2)	78 (35)	69 (30.9)	22 (9.9)	7 (3.1)
4. Provide multiple opportunities for ELs to process content in group contexts, including with their English-speaking monolingual or proficient bilingual peers (e.g., peer work assignments, whole-class discussions).	2.99 1.042	11 (4.9)	61 (27.4)	65 (29.1)	48 (21.5)	16 (7.2)
5. Include language objectives alongside content objectives in planning lessons.	2.49 1.059	34 (15.2)	78 (35)	54 (24.2)	26 (11.7)	9 (4.0)
6. Provide ELs with oral and visual support (e.g., think-alouds, word picture cards) to produce language orally and in writing.	3.27 .995	4 (1.8)	44 (19.7)	70 (31.4)	59 (26.5)	23 (10.3)
7. Provide alternative forms of assessments for ELs to gauge their comprehension and production of text.	2.83 1.013	12 (5.4)	72 (32.3)	65 (29.1)	36 (16.1)	13 (5.8)
8. Implement various strategies to differentiate instruction for ELs' success in language and literacy development.	2.98 1.012	10 (4.5)	58 (26)	74 (33.2)	42 (18.8)	16 (7.2)
9. Supplement curriculum and textbook materials with other sources to aid EL learning in content areas.	2.78 1.044	16 (7.2)	74 (33.2)	63 (28.3)	33 (14.8)	14 (6.3)
10. Teach ELs new vocabulary in context, both intentionally and incidentally.	3.00 1.035	10 (4.5)	59 (26.5)	71 (31.8)	42 (18.8)	18 (8.1)

Rating Scale Item	Frequencies					
	M SD	*Less Skilled* No. (%)	*Somewhat Skilled* No. (%)	*Skilled* No. (%)	*Very Skilled* No. (%)	*Highly Skilled* No. (%)
11. Apply various methods to incorporate different interactional and task engagement styles that ELs might bring from their cultural backgrounds into classroom work.	2.56 1.052	29 (13)	75 (33.6)	59 (26.5)	26 (11.7)	10 (4.5)
12. Develop higher order thinking skills for ELs at any English proficiency level through questioning and elicitation techniques.	2.47 .954	27 (12.1)	81 (36.3)	60 (26.9)	22 (2.2)	5 (2.2)
13. Encourage critical thinking by acknowledging ELs' diverse cultural experiences (e.g., arguments from authority may be highly valued in some cultures).	2.39 1.009	39 (17.5)	75 (33.6)	53 (23.8)	25 (11.2)	4 (1.8)
14. Help ELs to understand discipline-specific concepts within a content area through the use of graphic organizers (e.g., semantic maps).	2.70 .977	18 (8.1)	68 (30.5)	73 (32.7)	27 (12.1)	9 (4)
15. Adapt texts (e.g., paraphrasing unfamiliar expressions) to make content-specific concepts accessible to ELs.	2.71 1.019	20 (9)	68 (30.5)	67 (30)	31 (13.9)	10 (4.5)
16. Explicitly explain to ELs genres (e.g., expository, narrative) that are applicable to a specific content area.	2.58 .964	24 (10.8)	71 (31.8)	67 (30)	27 (12.1)	5 (2.2)
17. Build on ELs' knowledge of cognates between English and their home language.	2.24 1.041	52 (23.3)	73 (32.7)	46 (20.6)	17 (7.6)	6 (2.7)
18. Distinguish between core content vocabulary and common everyday vocabulary.	2.78 1.002	14 (6.3)	71 (31.8)	63 (28.3)	37 (16.6)	10 (4.5)
19. Pace oral communications and instructions appropriately for ELs' comprehension (e.g., repeat or clarify ideas, avoid or teach idioms and slang).	2.89 1.042	14 (6.3)	60 (26.9)	69 (30.9)	37 (16.6)	15 (6.7)
20. Develop ELs' metalinguistic awareness of the English language (i.e., ability to distinguish between literal and implied meanings).	2.47 1.009	31 (13.9)	77 (34.5)	57 (25.6)	22 (9.9)	7 (3.1)
21. Design written and oral activities to provide ELs at any English proficiency level with the opportunities to express their ideas and perspectives.	2.58 1.011	26 (11.7)	72 (32.3)	61 (27.4)	28 (12.6)	7 (3.1)
22. Help ELs identify areas of difficulty in academic vocabulary during in-class discussions of textbook passages.	2.63 1.026	26 (11.7)	64 (28.7)	69 (30.9)	26 (11.7)	9 (4.0)

Rating Scale Item	Frequencies					
	M SD	*Less Skilled* No. (%)	*Somewhat Skilled* No. (%)	*Skilled* No. (%)	*Very Skilled* No. (%)	*Highly Skilled* No. (%)
23. Help ELs decode meaning from highly abstract and culturally embedded phrases and sentences by encouraging them to infer meaning from context.	2.38 .932	33 (14.8)	79 (35.4)	60 (26.9)	19 (8.5)	3 (1.3)
24. Transform textbook content into meaningful chunks for ELs (e.g., highlight key concepts).	2.64 .952	18 (8.1)	75 (33.6)	66 (29.6)	29 (13)	6 (2.7)
25. Scaffold ELs' ability to rephrase or paraphrase academic language in their own words.	2.68 .973	15 (6.7)	79 (35.4)	63 (28.3)	28 (12.6)	9 (4.0)
26. Know that the discourse of academic texts in content areas includes the use of passive voice in describing events or explaining cause and effect.	2.36 .973	36 (16.1)	81 (36.3)	52 (23.3)	21 (9.4)	4 (1.8)
27. Making explicit the transition from everyday oral language to more technical language or register of the content area.	2.51 .990	28 (12.6)	74 (33.2)	61 (27.4)	24 (10.8)	6 (2.7)

Appendix 3

IRB Approval Letter

Office of Research Compliance
115 Ramsay Hall, basement
Auburn University, AL 36849

AUBURN
UNIVERSITY

Telephone: 334-844-5966
Fax: 334-844-4391
IRBadmin@auburn.edu
IRBsubmit@auburn.edu

February 9, 2015

MEMORANDUM TO: Ms. Heidi Goertzen
Department of Educational Foundations, Leadership, and Technology

PROTOCOL TITLE: "Examining Teacher Qualifications, Characteristics, and Practices Related to English Learners in the State of Alabama: An Exploratory Study on Teacher Quality"

IRB FILE NO.: 14-549 EX 1412

APPROVAL DATE: December 15, 2014
EXPIRATION DATE: December 14, 2017

Your protocol was approved as "Exempt" by the IRB under 45 CFR 46.101 (b) (2):

> Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless:
> (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and
> (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

Note the following:

1. CONSENTS AND/OR INFORMATION LETTERS: Only use documents that have been approved by the IRB with an approval stamp or approval information added.
2. RECORDS: Keep this and all protocol approval documents in your files. Please reference the complete protocol number in any correspondence.
3. MODIFICATIONS: You must request approval of any changes to your protocol before implementation. Some changes may affect the assigned review category.
4. RENEWAL: Your protocol will expire in three (3) years. Submit a renewal a month before expiration. If your protocol expires and is administratively closed, you will have to submit a new protocol.
5. CLOSING THE PROTOCOL: When your study is complete, please notify the Office of Research Compliance, Human Subjects.

If you have any questions concerning this Board action, please contact the Office of Research Compliance.

Bernie R. Olin, Phar.D.
Chair of the Institutional Review Board #2
for the Use of Human Subjects in Research

cc: Lisa Kensler

Appendix 4

Information Letter

INFORMATION LETTER

for a Research Study entitled *"Examining Teacher Qualifications, Characteristics, and Practices Related to English Learners in the State of Alabama: An Exploratory Study"*

You are invited to participate in a research study that investigates teacher practices for English learners. Heidi Goertzen is conducting this study under the direction of Dr. Lisa Kensler, Associate Professor in the Auburn University Department of Educational Foundations, Leadership, and Technology. You were selected as a possible participant because you are a general classroom teacher in a school that serves 20 or more ELs, and you are age 19 or older.

Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey. The only risk or discomfort associated with this study is the time needed to complete the survey. To minimize this discomfort, your total time commitment to complete the survey will be approximately 10 minutes.

If you participate in this study, you will receive the results of the study and will be contributing to the knowledge base about teacher practices for ELs. If you are interested in receiving the results of my study, you will have an opportunity at the end of the survey to follow a link to a new survey and provide your email address. I will email you aggregated results of the study upon completion. There is no cost to participate in this study other than the time it takes to complete the survey.

Data you provide in connection with this study will remain anonymous. Information obtained through your participation will be used to complete the doctoral dissertation of Heidi Goertzen, may be published in a professional journal, and may be presented at a professional conference.

If you have any questions, please contact me at htg0003@auburn.edu or my advisor, Dr. Lisa Kensler at lak0008@auburn.edu. If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334)-844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION ABOVE, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, PLEASE CLICK ON THE LINK BELOW. YOU MAY PRINT A COPY OF THIS LETTER TO KEEP.

The Auburn University Institutional Review Board has approved this document for use from December 15, 2014 to December 14, 2017. Protocol #14-549 EX 1412

Heidi Goertzen, December 5, 2014
Investigator **Date**

You can access the survey at this link:
https://auburn.qualtrics.com/SE/?SID=SV_cSiu7iOZOwWyFxj

Thank you for your consideration.

Appendix 5

Survey Reminder Letter

Good morning,

I hope all is well as we near a much anticipated spring break! A couple of weeks ago I sent you an email regarding your interest to invite teachers to participate in survey research on teacher practices for English learners. Before I close the survey on April 3rd, I wanted to reach out once more to invite their voluntary participation.

If interested, please share the survey link with them located at the end of the attached consent information letter.
With much appreciation,
Heidi