2018

A Quantitative Investigation of the Relationship Between English Language Assessments and Academic Performance of Long-Term ELLs

Yesmi Rios
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

Yesmi Rios

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Peter Lownds, Committee Chairperson, Education Faculty
Dr. Andrew Thomas, Committee Member, Education Faculty
Dr. Andrea Wilson, University Reviewer, Education Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2018

Abstract

A Quantitative Investigation of the Relationship Between English Language Assessments and Academic Performance of Long-Term ELLs

by

Yesmi Rios

MA, San Francisco State University, 2001

MA, Harvard University, 1997

BAS, University of California Davis, 1996

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Policy, Leadership, and Change

Walden University

May 2018

Abstract

Research shows academic literacy is a challenge for students classified as Long-Term English Language Learners (LTELLs). In the pseudonymous Windy Desert School District (WDSD), there are 17,365 students classified as LTELLs. Of these students, the majority are falling short of English academic literacy goals on the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) test and 67% do not graduate from high school. This quantitative study examined the predictive relationship between ACCESS English language proficiency subscale scores in the language domains of speaking, listening, reading, and writing and course semester grades in English 9, English 10, and English 11. This longitudinal study, informed by theorists Cummins and Krashen, followed a cohort of 718 Grade 9 students for 3 years (2012–2015). Of the 718, only 161 participant data sets were valid for the final ordinal logistic regression analysis. ACCESS subscale scores in speaking, listening, reading, and writing comprised the predictor variables and English course semester grades comprised the criterion variables. Results revealed that LTELLs' ACCESS subscale scores in listening, reading, and writing were significant predictors of their English course grades whereas speaking scores were not. For each predictor variable, a 1-unit increase in the predictor decreased the likelihood of receiving a lower grade in the course. Social change can result from the WDSD using ACCESS results to create and implement effective instructional programs that develop LTELLs' proficiency in the language domains found significant in predicting their academic grades, thereby increasing their language proficiency, academic grades, and graduation rates over time.

A Quantitative Investigation of the Relationship Between English Language Assessments and Academic Performance of Long-Term ELLs

by

Yesmi Rios

MA, San Francisco State University, 2001

MA, Harvard University, 1997

BAS, University of California Davis, 1996

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Policy, Leadership, and Change

Walden University

May 2018

Dedication

To my mother, Maria Fernandez de Rios, for being loving, strong, smart, honest, dedicated, progressive, innovative, encouraging, and motivating. We accomplished it mama! Your daughter is a doctor! Thank you. I love you and miss you.

Acknowledgments

My love and admiration goes out especially to my husband, Victor, and son, Little Vic, who supported and endured the long journey. Thank you for your love, patience, and care. You are both extraordinary human beings and I love you for it.

To my chair, Dr. Peter Lownds, thank you for your enthusiasm, your positive energy, and your unwavering support. Your willingness to mentor, encourage, and elevate my spirits made this journey endurable. Paulo Freire stated "Reading is not walking on the words; it's grasping the soul of them." Your passion for words and human interaction makes you brilliantly distinct. Ciao, professor. We made it.

To my methodologist, Dr. Andrew Thomas, thank you for adopting me. Your encouraging remarks, your professionalism, your expertise, and your down-to-earth personality made all the difference. You are an inspiring example of what a professor should be.

Thank you Dr. Cheryl Keen, Dr. Andrea Wilson, Dr. Debbra Lindo, Drs. Brenda and Brian Daw, Dr. Nancy Alamo, Tracy Clark, Tom Bacon ("Captain America"), Mr. Brignac, Martha and Michael Morales, Michael Bear, Perla Carbajal, and Penny Nicholson for the time and knowledge you graciously offered.

My love and appreciation goes out to my sister who consistently inspired me with words, jokes, and pictures of my adorable nephews and to my four brothers, whose competitive spirit has always inspired me. I also thank my father for his jokes, his presence, and for making education a priority.

For many others, too numerous to mention here, thank you for championing me.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Twenty percent of public school students in the United States are the daughters and sons of immigrants who speak a language other than English at home (Gándara & Hopkins, 2010). Many of these children fall short of *fluent* in English on their state's language proficiency assessment scales and are relegated to the status of English Language Learners (ELLs). Researchers have estimated that 5 million ELLs are currently enrolled in elementary, middle, and high schools throughout the United States (Heritage, Walqui, & Linquanti, 2015).

Within the large mass of ELLs, there exists a subgroup of students referred to as *Long- Term English language learners* (LTELLs) who have attended U.S. schools for 5 or more years without attaining sufficient English language proficiency (Windy Desert School District, 2010). Researchers Menken, Kleyn, and Chae (2012) found that many of these students speak colloquial English but read and write far below grade level. This means that LTELLs may sound like native speakers but have limited proficiency when reading or writing English. Due to their oral fluency, LTELLs are often mainstreamed into regular classes where they receive inconsistent pedagogical support (Freeman, 2002; Menken et al., 2012; Olsen, 2014). When LTELLs perform poorly in school, they are often referred to special education classes under the assumption that their academic difficulties are cognitive rather than linguistic (Freeman & Freeman, 2009).

There are many LTELLs amassing in middle and high schools across the United States (Olsen, 2010). Generally, LTELLs are often neglected by teachers who have not

been trained to meet their unique educational needs (Olsen, 2010). Olsen (2010) was the first researcher to reveal to educators and policymakers the educational challenges that prevent LTELLs from acquiring the literacy skills they need to successfully graduate from high school (Menken et al., 2012; Olsen, 2010). This problem is especially pronounced in public schools located in Los Angeles, Chicago, Miami, Las Vegas, and New York, where 35–50% of students whose primary language is not English are classified as LTELLs (Menken et al., 2012; New York City Department of Education, Office of English Language Learners, 2008; Olsen, 2010). The adoption of college- and career-ready standards by U.S. public schools presents additional challenges for these students (Bailey & Carroll, 2015). LTELLs are often unprepared for standards that encourage deep, transferable content learning and analytical practices (Bailey & Carroll, 2015). Thus, they fail required course content and drop out of high school in disproportionately high numbers (Heritage et al., 2015; Mathews, 2007).

Nationwide information on U.S. LTELLs' English proficiency is derived from English language assessments that the Every Student Succeeds Act (ESSA)—which replaced the No Child Left Behind (NCLB) Act in 2017—requires states to administer annually to all ELLs. These tests are used to assess both LTELLs' social and academic language competency. The Assessing Comprehension and Communication in English State-to-State (ACCESS) for ELLs was the English language Proficiency Assessment (ELPA) used in this study. ACCESS was developed by the World-Class Instructional Design and Assessment (WIDA) Consortium and currently implemented in 39 states.

ACCESS is used to assess students' language competence in the domains of speaking, listening, reading, and writing across five WIDA English language development standards represented as *Social and Instructional language*, *The language of Language Arts*, *The language of Mathematics*, *The language of Science*, and *The language of Social Studies* (WIDA, 2013). An important element of these five standards is their proven alignment with academic content standards that guide instructional classroom practices delineated by each state (Cook, 2007; WIDA, 2013).

In the pseudonymous Windy Desert School District (WDSD), the study site school district, school leaders use ACCESS oral proficiency scores to determine LTELLs' course placement (Abedi, 2007; U.S. Department of Education, 2013). The WDSD educational leaders assume that LTELLs' oral fluency on ACCESS is an accurate predictor of LTELLs' ability to understand course content and perform well in classroom speaking, listening, reading, and writing tasks (Gandy, 2013). However, being able to speak English proficiently may not mean a student can use academic language well enough to succeed in courses. If a student is placed incorrectly based solely on his or her oral fluency, that student may fail to progress academically.

The aim of this study was to explore the predictive relationship between ACCESS scores and LTELLs' academic performance in high school English courses, specifically English 9, English 10, and English 11. Researchers have acknowledged that well-developed proficiency in the areas of speaking, listening, reading, and writing are essential requirements for students who want to communicate effectively in everyday

situations and master academic content matter (Callahan, 2005; Cummins, 1984; DiCerbo, Anstrom, Baker, & Rivera, 2014; Olsen, 2014; Snow & Uccelli, 2009). I selected high school English courses because they are considered the most effective at engaging students in a curriculum focused on specific linguistic features, such as grammar and vocabulary, which inform the language domains of listening, speaking, reading, and writing across a content spectrum that includes mathematics, science, and history courses (Bailey & Heritage, 2008; Gottlieb, Katz, & Ernst-Slavit, 2009; Schleppegrell, 2004). The grades acquired by LTELLs in English classes, therefore, serve as a basis for assessing their proficiency in several core subjects. Moreover, because the English Language Arts (ELA) academic content standards that guide the instructional practices of WDSD English 9, English 10, and English 11 teachers have been correlated to the ACCESS test, these assessment scores should be able to predict students' academic achievement in high school English courses (Chi, Garcia, Surber, & Trautman, 2011).

Researchers have investigated the relationships between ACCESS scores and English scores on standardized state assessments (Crane, Barrat, & Huang, 2011; Forte, Perie, & Paek, 2012; Grant, Cook, & Phakiti, 2011). Although standardized tests are important, some researchers argue that teacher-assigned grades have a more significant consequence, given the role they play in course placement, high school graduation, and college or career opportunities (Cornwell, Mustard, & Van Parys, 2008). In this study, course semester grades represented students' academic achievement in English courses. My intent was to add to the existing research by exploring if, and to what extent, English

proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment can predict semester course grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students enrolled in the WDSD.

In the following sections, I provide the background on the study site school district, define the research problem, and provide the purpose or justification for investigating the problem. I state the questions guiding my study, describe the theoretical framework, and briefly outline the nature of this study. Definitions, assumptions, scope and delimitations, limitations, significance, and an overall summary are also included.

Background

The WDSD, located in a Sun Belt metropolis, ranks sixth in the nation in the number of ELLs enrolled. More than 30% of the district's 325,032 students come from households where the predominant language is other than English. Of these students, more than 59,234 are considered ELLs. According to 2017 data from the WDSD, 17,365 of these ELLs are LTELLs. In 2016–2017, only 2% of LTELLs obtained English language proficiency. The dropout rate among LTELLs that same year was 66% (WDSD, 2017). Slama (2012) demonstrated that a variety of factors may contribute to students dropping out: academic performance, grade retention, school engagement, and the inability to pass tests dependent on academic language (Bailey & Carroll, 2015; Olsen, 2014; Suárez-Orozco & Suárez-Orozco, 2001; Thomas, 2005; Wong-Fillmore & Snow, 2000). Olsen (2014) found that adolescent LTELLs who failed to acquire adequate academic language skills were more likely to drop out of school. Failing to graduate from

high school severely limits job opportunities and college matriculation, consigning these students to a life of poverty (Slama, 2012).

Concerned with the low language proficiency and high dropout rates of LTELLs, the WDSD asked educators, parents, students, and community stakeholders to collaborate in a careful examination of the problems they confront. This effort included direct observation of students in hundreds of first- through 12-grade classrooms at multiple schools. School district leaders analyzed a variety of data: LTELLs' linguistic and academic performance, syntheses of dozens of focus groups and interviews, and thousands of surveys involving students, parents, community stakeholders, teachers, principals, and district administrators. Community feedback was also welcomed in several public forums. The systemic review of the WDSD's current educational policies and practices demonstrated that despite the implementation of new content standards, relatively few LTELLs are placed in courses where they receive instruction that offers them grade-appropriate, academically rigorous opportunities for language learning (WDSD, 2016).

In the WDSD, LTELLs' abilities to understand, speak, read, and write English is assessed annually using the ACCESS assessment test (WIDA Consortium, 2015). Every student classified as an ELL, including ELL special education students, must complete the test. The main purpose is to measure the annual progress of ELLs toward the attainment of English language proficiency. ACCESS scores identify six levels: 1-Entering, 2-Beginning, 3-Developing, 4-Expanding, 5-Bridging, and 6-Reaching. The

levels outline the progression of ELL language development and delineate classroom performance in speaking, listening, reading, and writing tasks (Gottleib, Cranley, & Cammilleri, 2007). The WDSD data (2016) demonstrate that most LTELLs have an overall classification of *developing* (Level 3) or *expanding* (Level 4) on ACCESS and commonly demonstrate Level 5 or 6 proficiency only on the speaking component of the test. Many school administrators and teachers assume that this oral fluency on ACCESS validates LTELLs' proficiency in academic language (Olsen, 2014). Thus, these administrators place LTELLs in mainstream classes where they are taught the regular, core curriculum without compensatory language support, accommodations, or strategies to help them master course content (Olsen, 2010).

Researchers have demonstrated that placing ELLs in mainstream classes without adequate language support produces the worst achievement outcomes (Lindholm-Leary & Genesee, 2010; Olsen, 2010; Thomas & Collier, 2002). High school ELLs placed in mainstream programs receive lower grades than their peers who are placed in ELL-only programs (Lindholm-Leary & Genesee, 2010). The WDSD data from the 2013–2014 state assessments in ELA confirm these findings. For example, the 2013–2014 data indicated that 20.1% of high school ELLs received *proficient* scores on the ELA assessments, whereas ELA proficiency was achieved by 64.2% of high school students whose first language was English (WDSD, 2016).

Olsen (2014) found LTELLs benefitted from being (a) placed in rigorous grade-level content courses, (b) offered courses in primary language literacy and academic

language development, and (c) monitored by established systems regarding their progress. Zwiers (2008) observed LTELLs benefitted from being in language-rich environments where they were provided the opportunity to acquire language through meaningful interactions with others. The WDSD data (2015) showed that LTELLs who attain English language proficiency match or exceed the academic performance of non-ELL students (WDSD, 2016). Lindholm-Leary and Genesee (2010) and Thomas and Collier (2002) found that high school students who had participated in ELL programs were more likely to score at grade level and less likely to drop out than those who were prematurely moved to mainstream classes.

The clear conclusion emerging from the research on LTELLs is that English proficiency is required for academic success. Considering the important role that English course content plays in the development of students' language and literacy skills, understanding the predictive relationship between the English language proficiency of LTELLs and their level of mastery in English course content is especially critical (Snow & Kim, 2007). Schleppegrell (2007) argued that language and learning content should not be separated. Hakuta (2014) noted that teachers should be engaged in the constant development of their students' academic English as well as in the subject matter required by the curriculum. Researchers have confirmed that ELL performance on English state academic tests is influenced by their English language proficiency (Crane et al., 2011; Forte et al., 2012; Grant et al., 2011). My intent was to add to the existing research by examining if, and to what extent, English proficiency subscale scores in speaking,

listening, reading, and writing on the ACCESS assessment can predict course semester grades in English 9, English 10, and English 11 for the 2012–2015 cohort of ninth-grade LTELL students. The results from this study may affect the manner in which the WDSD and other districts within the state use assessment results to create more equitable learning opportunities for LTELLs in the form of services, programs, and support.

Problem Statement

Federal law mandates that public school districts must address the academic needs of LTELLs (U.S. Department of Education, 2016). However, the WDSD fails to meet these needs. In 2016–2017, only 2% of LTELLs' obtained English language proficiency and only 33% graduated from high school. One reason for this failure may be that school administrators and teachers assume that LTELLs' oral fluency on ACCESS indicates proficiency in academic language (Olsen, 2014). Consequently, these educators use ACCESS oral proficiency scores for LTELL course placement. However, being able to speak English proficiently may not mean a student can use academic language well enough to succeed in courses. If a student is placed incorrectly based solely on his or her oral fluency, that student may fail to progress academically. Although LTELLs often demonstrate proficiency in speaking, data show they do not obtain proficiency on the ACCESS test overall and many of these students fail to graduate (WDSD, 2016). Through this study, I addressed the problem of using ACCESS oral proficiency scores to predict LTELL academic success, as measured by English course grades.

Research has demonstrated that English proficiency in the areas of speaking, listening, reading, and writing can raise LTELLs' academic grades (Aina, Ogundele, & Olanipekun, 2013; Fakeye, 2014; Ghenghesh, 2015; Kumar, 2014; Sahragard & Baharloo, 2009). Researchers have also explored the predictive value of English proficiency assessments to earmark students for interventions and to determine specific instructional strategies on high school exit exams, end-of-course exams, and norm-referenced tests (Georgia Department of Education, 2008; Gewertz, 2007; Parker, Louie, & O'Dwyer, 2009; Wakeman, 2013). However, the possibility that a predictive relationship exists between LTELLs' ACCESS subscale scores and their teacher-assigned letter grades in high school English courses had not been examined.

This quantitative study focused on a cohort of WDSD LTELL students and evaluated their achievement over 3 years in English 9, English 10, and English 11. The focus on English courses was predicated on the fact that the ACCESS assessment measures English language proficiency in the domains of speaking, listening, reading, and writing and that the high school English curricula are based on linguistic features, such as grammar and vocabulary, which relate to these language domains (Bailey & Heritage, 2008; Gottlieb et al., 2009; Schleppegrell, 2004). Because English classes offer LTELLs the opportunity to attain proficiency in all four of the language domains assessed by ACCESS, it follows that performance in English and ACCESS scores should have a positive, linear relationship. An outcome of this study was a deeper understanding of the relationship between ACCESS subscale scores and student English course grades.

The results also offer insight into whether oral proficiency scores are used appropriately in the WDSD and elsewhere, and whether other ACCESS subscales (listening, reading and writing) are significant predictors of academic success.

Purpose of the Study

The purpose of this quantitative study was to examine if, and to what extent, English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment can predict semester course grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students. ACCESS subscale scores comprised the predictor variables and English course semester grades comprised the criterion variables. ACCESS subscale scale scores were calculated and reported for the language domains of speaking, listening, reading, and writing of the ACCESS test. The lowest possible scale score was 100 and the highest was 600. Scale scores were used in the study because they are continuous and independent variables that made it possible to compare test scores across grades and tiers, within any of the four language domains.

Letter grades A–F were used to assess LTELLs' academic performance in English 9, English 10, and English 11 courses. Letter grades were operationalized as ordinal ranked data, 1–5. The decision to use a cohort of ninth-grade LTELLs was made based on the implications that the results of a predictive study might have to address the needs of future incoming ninth-grade LTELLs who remain enrolled in the WDSD. By examining the predictive value of the ACCESS assessment scores on LTELLs' English course semester grades, I provided an increased understanding of how the language

domains affect LTELLs' English grades and identified potential considerations for instructional practices that best support their needs.

Research Questions and Hypotheses

The following research questions and hypotheses guided this study. I addressed each question using ordinal logistic regression analysis.

RQ1. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2012–2013 assessment predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$1a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment do not predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a1a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ2. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2012–2013 assessment predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$2a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–-2013 assessment do not predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a2a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ3. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2013–2014 assessment predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$3a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment do not predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a3a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ4. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2013–2014 assessment predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$4a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–-2014 assessment do not predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a4a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ5. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2014–2015 assessment predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$5a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment do not predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a5a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ6. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2014–2015 assessment predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$6a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment do not predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a6a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

Theoretical Framework

The theoretical framework for this study was based on theories by Cummins (1979) and Krashen (1982). Cummins' theory of second language acquisition explains the challenges students face as they acquire fluency in conversational and academic English. Krashen's monitoring hypothesis explains the psychological and cognitive processes of second language acquisition, which I used to describe how educators can harness this knowledge to advance LTELL language development. In the following chapter, I provide a more in-depth examination of the two theories.

Thomas and Collier (1997) defined *success* for second language learners as "eventually reaching full educational parity with native-English speakers in all school content subjects (not just in English) after a period of 5–6 years" (p. 7). Researchers have found that LTELLs who have not reached English language proficiency after 6 years tend to perform poorly in academic classes (Bailey & Carroll, 2015; Olsen, 2014; Slama, 2012; Stella, Batalova, & Fix, 2012). If these theorists and researchers are correct, students need to be exposed to high levels of academic language in order to achieve

academic success. In Chapter 2, I use the work of Cummins and Krashen to explain certain impediments LTELLs encounter when internalizing and applying acquired English language skills in academic settings. I also detail the type of academic language LTELLs need to acquire to experience sustained academic achievement.

Nature of the Study

Because ACCESS scores and academic achievement in English classes are quantifiable variables, the nature of my study was quantitative and followed a nonexperimental design. Nonexperimental research does not involve manipulation of the independent variable and can be used to assess the predictive relationship among variables. A correlation is defined as the measurement of the relationship between two variables (Lomax, 2013). For this study, I sought to investigate relationships between two variables; a correlational design is appropriate for this type of study (Stevens, 2009). Specifically, I assessed the predictive relationship between students' English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment and their course semester letter grades in English 9, English 10, and English 11. Because I did not control, manipulate, or alter the predictor variable or subjects, a nonexperimental design was appropriate (Pagano, 2010). I conducted ordinal logistic regression analyses to assess the presence of these predictive relationships.

The study cohort included the entire group of ninth-grade students who were categorized as LTELLs during the 2012–2013 school year and then persisted in the WDSD for 3 consecutive school years. The study participants comprised students from

53 high schools within the WDSD. The original database file included a sample size of 718 LTELL students; however, students with incomplete data sets (n = 218, 30%) were not included in the analysis. A complete data set consisted of speaking, listening, reading, and writing ACCESS scores on successive tests administered in the first semester of the 2012, 2013, and 2014 school years and fall and spring semester letter grades in English 9, English 10, and English 11 courses from the 2012–2013 to the 2014–2015 school years.

Furthermore, students identified as special education students (n = 218, 30%) were not included in the analysis. Students enrolled in one or more Honors English 9, English 10, or English 11 course semesters (n = 65, 9%) or who repeated one or more English 9, English 10, or English 11 course semesters (n = 24, 3%) were also excluded. Additionally, 32 outliers were assessed and subsequently removed from the data set. Unexpectedly, 78% (n = 557) of my original sample size was excluded from the analysis. The final analytic sample consisted of 161 LTELLs.

Archival data on students' ACCESS scores for each year between 2012 and 2015 were drawn from the district's Infinite Campus database. I selected this time frame because the WDSD first implemented the ACCESS assessment during the 2012–2013 school year. Thus, the period between 2012 and 2015 represented the only 3 years of complete data available. Ordinal logistic regression analysis was employed to assess the predictive relationship between the students' English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment and their subsequent

fall and spring course semester grades in high school English 9, English 10, and English 11.

Definitions

Academic achievement. Students' success in course content measured through teacher-assigned letter grades (Roksa & Potter, 2011).

Academic language. The oral and written language used in school settings that is characterized by linguistic features such as discourse, grammar, and vocabulary and used to describe abstract concepts, higher order thinking, and complex ideas (DiCerbo et al., 2014; Dutro & Moran, 2003).

Assessing Comprehension and Communication in English State-to-State (ACCESS) for English Language Learners (ELLs). An English language proficiency assessment given to K–12 grade ELL students that measures their attainment of English proficiency in the language domains of speaking, listening, reading, and writing (WIDA Consortium, 2015).

Basic interpersonal communication skills (BICS). The level of conversational fluency demonstrated by second language learners within approximately two years of being exposed to a second language (Cummins, 1979).

Cognitive academic language proficiency (CALP). The oral and written academic language of the classroom that is believed to take ELLs 5–7 years to acquire (Cummins, 2000).

English Language Learner (ELL). An individual not yet proficient in the English language. In the United States, some educators prefer the term, *emergent bilinguals*, instead of ELL (Cheatham, Jimenez-Silva, Wodrich, & Kasai, 2014).

English language proficiency assessment (ELPA). An annual test used to measure the degree to which ELLs are making progress towards acquiring English language proficiency (Forte et al., 2012).

First language (L1). A language that a student first learned or was exposed to from birth (Cummins & Hornberger, 2008).

Language acquisition. The process of learning the units of a language, as well as how to construct them, to communicate verbally (Saffran, Senghas, & Trueswell, 2001).

Limited English Proficient (LEP). A student not yet proficient in the English language (Cheatham et al., 2014). This term is synonymous with ELL.

Long-Term English Language Learner (LTELL). An ELL student who has attended U.S. schools for 5 or more years without acquiring English language proficiency (WDSD, 2011).

Second language (L2). The language learned after acquiring the first language (Cummins & Hornberger, 2008).

Assumptions

A few assumptions were inherent to this study. First, I assumed that all data had been accurately recorded and archived, and that the data set contained complete information for the study population. This assumption was based on the fact that the

WDSD had standardized instruments, administration of data collection procedures, and ongoing training and data screenings to ensure that all archived data were accurately recorded and complete.

This study also rested on the assumption that ACCESS effectively measures language proficiency of LTELLs and accurately categorizes students into proficiency levels. This assumption was based on information provided by WIDA validation studies that demonstrated that ACCESS appropriately measures test takers' different levels of social and academic English proficiency and classifies them into appropriate English language proficiency levels (WIDA, 2015). I assumed ACCESS was administered and evaluated with the same consistency during the 3 academic years that were the focus of my study (2012–2013, 2013–2014, 2014–2015). The WDSD administered ACCESS to more than 68,000 students each year. School personnel at the different school locations administered the listening, reading, and writing sections. The speaking section of ACCESS, which was required to be administered individually to each student, was administered by one of the 54 district testers who was trained to administer and score the speaking section. The WDSD had standardized instruments, administration of data collection procedures, and required trainings to help control the instrumentation threat.

As it pertains to letter grades, I assumed teachers followed district regulations when assessing students and I assumed the professional learning communities were successful in ensuring grading consistency among WDSD high school English teachers. Data analysis rested on a standard set of assumptions that underlie ordinal logistic

regression analysis. I assumed that individual student observations were independent. My intent was to assess the predictive relationship between the ACCESS scores and student grades in their English courses, a relationship that educators previously assumed was present. For regression, I assumed that only student grades (course marks) were random variables. For hypothesis testing, I assumed that student grades were distributed according to the normal distribution and the variability of student grades was the same at each value of the relevant ACCESS score. For the ordinal logistic regression analysis, I assumed the relationship between the two variables was linear. A more detailed discussion of these assumptions is included in Chapter 3. In Chapter 4, I describe the procedures used for testing the statistical assumptions.

Scope and Delimitations

Simon (2011) defined *delimitations* as those characteristics that limit the scope and define the boundaries of a study. The scope of this study was limited by space and time, as data were drawn only for LTELL students who attended one of the 53 high schools in the WDSD from 2012 to 2015; the study did not include elementary or middle school LTELL students. The WDSD initiatives and needs influenced my decision to focus on high school LTELL students. The decision to use a cohort of ninth-grade LTELLs was made based on the implications that the results of a predictive study might have to address the needs of future incoming ninth-grade LTELLs who remain enrolled in the WDSD. The study was further delimited by those LTELLs who were present in the

district all 3 years and had all ACCESS scores; only these students were included in the data analysis.

The lens through which I viewed language acquisition and student achievement may have been a delimiting factor that affected the scope of my understanding. I could have modeled second language learning on the sociocultural framework posited by Lev Vygotsky (1978), who argued that the use of language and interaction in a social environment extends the cognitive abilities of an individual. Vygotsky also believed language was the actuator of learning (Wertsch & Tulviste, 1992). However, I chose not to use Vygotsky's theory because his framework was based on how the use of language affects learning, rather than how learning is influenced by the acquisition of language. I chose Cummins's (1979) and Krashen's (1982) theories of second language acquisition instead because the theories help to explain the psychological and cognitive processes LTELLs use to acquire second language proficiency.

Limitations

This study had a few limitations that I addressed. First, because this study was delimited by a single school district, the sample was a unique, nonprobability convenience sample, rather than a random sample. The lack of a random sample denotes that results may not be generalizable beyond the specific population from which the sample was drawn. This limitation applies to all studies that are not randomized controlled trials. However, external validity was enhanced because 39 states that are part of the WIDA consortia have adopted ACCESS. ACCESS was chosen as the focus of this

investigation because it is the most widely used English language assessment in U.S. schools, as well as the assessment used in the WDSD (WIDA, 2016). However, ACCESS is not the only ELPA used in the United States (WIDA, 2016). Districts in states that are part of the English Language Proficiency Assessment for the 21st Century (ELPA21) consortia use two assessments that measure ELL proficiency in English, mathematics, and science (ATLAS, 2015). Several states that are not part of the WIDA or ELPA21 consortia use the English Language Development Assessment (ELDA) or the Language Assessment Scales Links assessments to measure students' English language proficiency in the four domains of reading, writing, speaking, and listening (ATLAS, 2015). In addition, several states, including Arizona, California, Florida, New York, Ohio, Texas, and Washington, have developed their own language proficiency assessments (ATLAS, 2015). Thus, any findings from this study that show a connection between ELPA scores and grades should not be interpreted as possibly holding true in states that do not use the ACCESS assessment.

I recognized that there are LTELLs in different grades and levels of development and language acquisition. However, my study was limited to a cohort of ninth-grade LTELLs. The selection of a ninth-grade LTELL cohort allowed tracking of data for 3 school years and time for students to be reclassified as English language proficient. The study of a ninth-grade LTELL cohort also has implications for future incoming ninth-grade LTELLs who remain enrolled in the WDSD. I only analyzed ACCESS scores and English fall and spring course semester grades for these students.

An additional limitation to this study was that student characteristics, such as LTELL special education status, LTELL honor course status, or LTELL repeat course status, were not assessed. Subject characteristics threaten the design when the possibility exists that individual characteristics other than those already correlated can explain any relationships that are found (Fraenkel & Wallen, 2012). The disability level of the subjects classified as both LTELLs and special education students, for example, can be expected to relate to English language proficiency as well as to academic performance. The likelihood of such characteristics having an effect was high. Therefore, these groups of LTELLs were excluded from the design to eliminate this subject characteristic threat.

A final and major limitation of the study was the variability found within each WDSD high school as it pertained to teachers, instructional practices, and English course content that may have affected letter grades. This limitation was further exacerbated because each student's data set consisted of two course semester grades from three different English courses with distinct curriculums. To reduce the variability of their grades, WDSD English teachers met and developed common homework assignments, common assessments, and common grading practices for their content area. *Professional learning community* leaders from each high school content area attended district-wide trainings. The content leaders returned to their schools and shared information from district-wide trainings with their colleagues. Thus, the WDSD implemented measures to ensure that grades were valid and reliable indicators of student achievement by issuing district-wide regulations and sponsoring professional learning communities (WDSD,

2016). Other factors that can influence students' grades, such as their attendance rates, motivation, and socioeconomic status, did not fall within the scope of this investigation.

Significance

Addressing the academic needs of LTELLs is a provision included in the ESSA and a WDSD initiative. In the United States, 39 state education agencies have adopted ACCESS as their assessment tool to measure LTELL students' annual progress in the attainment of English language proficiency (WIDA, 2016). Educational leaders and policy makers assume that English language proficiency as measured by the ACCESS assessment is an accurate indicator of LTELL students' abilities to understand course content and perform well in classroom speaking, listening, reading, and writing tasks (Gandy, 2013). However, no known researchers have explored the predictive relationship between LTELL students' ACCESS scores and their academic performance in high school English courses. Thus, the aim of this study was to explore the predictive relationship between ACCESS scores and LTELL students' academic performance in high school English 9, English 10, and English 11 courses. English courses were the focus because they engage students in speaking, listening, reading, and writing tasks that have the potential to make the students effective readers, writers, speakers, and listeners in every content area (Common Core State Standards Initiative, n.d.). Moreover, the curricular standards guiding English 9, English 10 and English 11 teachers were correlated to the ACCESS test (Chi et al., 2011).

The results of this study may contribute to social change at the school, district, and state level. Based on the findings of this study, WDSD English teachers and school administrators can be proactive at identifying and implementing instructional programs that help LTELLs develop their English language proficiency in the language domain(s) that most affect their English grades. Such findings may ultimately affect the funding, staffing, and curricula of English programs that focus on increasing the academic language development, achievement scores, and graduation rates of LTELLs.

Summary

The dropout rate among LTELLs in the WDSD is 66% (WDSD, 2016). Their English language proficiency is assessed by the ACCESS exam for ELLs (WIDA, 2015). Thus, ACCESS plays a critical role in determining the course placement, curriculum, type of instruction, instructional materials, and language support these students receive (Abedi, 2007; U.S. Department of Education, 2013). A critical shortcoming of the ACCESS is that researchers have not validated its ability to predict students' levels of content knowledge, as measured by letter grades. Abedi (2007) indicated that commonly used language proficiency tests often assess a wide variety of skills that may not necessarily be useful in the attainment of academic success.

To comply with federal mandates to meet the academic needs of LTELLs, it is critical that language proficiency assessments be accurate indicators of students' abilities to comprehend course curricula. Thus, the aim of this study was to examine the extent that English proficiency subscale scores in speaking, listening, reading, and writing on

the ACCESS assessment can predict LTELLs' course semester grades in English 9, English 10, and English 11. The data may be used by stakeholders to hire more skillful and communicative teachers who will increase LTELL students' English language proficiency in the language domain(s) that most affect their English grades. The results of this study may also help teachers and school administrators throughout the state—who depend on ACCESS to monitor students' academic progress—become aware of variations in the efficacy of English courses designed with LTELLs in mind. Such findings may ultimately affect the funding, staffing, and curricula of English programs that focus on increasing the rate of academic language development, achievement scores, and graduation rates of LTELLs.

The aim of this chapter was to contextualize this study and orient the reader to the research problem, purpose, and questions. I discussed the study's theoretical framework, definitions, assumptions, scope, and delimitations. In Nature of the Study, I provided a preview of the methodology, which I further detail in Chapter 3. Finally, I explained the social significance of this research.

In Chapter 2, I examine the scholarly accounts of LTELL students' academic vicissitudes and explain the concept of English language proficiency. I discuss the theoretical basis of second language acquisition. I also provide a review and synthesis of current research on the predictive capacity of the ACCESS assessment and related concerns to expose a gap in the research. Finally, I describe the study methodology in Chapter 3.

Chapter 2: Literature Review

In a predominantly English-speaking society, literacy skills are critical to the academic success of LTELLs. Research suggests that many LTELLs require continuous training and support to develop literacy skills that will enable them to graduate from high school and succeed in college. A strong pedagogical focus on reading and writing may develop the academic language and literacy skills of these students (Heritage et al., 2015).

The NCLB Act of 2001 required states to develop and administer standards-based ELPAs that measured LTELLs' annual progress (Bailey & Carroll, 2015; Bailey & Wolf, 2012). Historically, ELPAs were designed to assess the social dimensions of language proficiency. With current theory, practice, and new legislative emphasis on the integration of language and curricular content, states must now establish, implement, and sustain high-quality instructional programs to ensure LTELLs become proficient in the academic as well as the social aspects of language acquisition (U.S. Department of Education, 2016). Test designers adapted to this change by creating ELPAs that reflected and assessed the academic language at the K–12 level. It is unclear, however, whether results from these new ELPAs can predict LTELLs' academic performance in high school content courses.

The focus of this quantitative study was to investigate to what extent English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment can predict course semester grades in English 9, English 10, and English 11

for the 2012–2015 cohort of ninth-grade LTELL students. ACCESS is a standardized assessment used by educational agencies in 39 states to gauge LTELLs' English language acquisition and literacy levels. Administrators and teachers within the WDSD assume that LTELLs' oral fluency on ACCESS indicates proficiency in academic language (Olsen, 2014). Consequently, these educators use ACCESS oral proficiency scores for LTELL course placement. However, being able to speak English proficiently may not mean a student can use academic language well enough to succeed in courses. If a student is placed incorrectly based solely on his or her oral fluency, that student may fail to progress academically. Therefore, it is important to understand the relationship between ACCESS subscale scores and student achievement, as measured by grades, to determine whether oral proficiency scores are used appropriately in the WDSD and elsewhere, and whether other subscales (listening, reading, and writing) are significant predictors of LTELL academic success.

In Chapter 2, I begin with an examination of scholarly accounts of LTELLs' academic vicissitudes followed by a brief history of the concept of English language proficiency, especially when it entails the acquisition of a second language. The chapter also includes a discussion of the theoretical basis of second language acquisition. Next, I review the current literature pertaining to ACCESS. Finally, I examine the research basis for using ELPAs to predict students' academic performance.

Search Strategy

I retrieved literature related to English language acquisition, English language assessments, ACCESS for ELLs, and academic English performance from books, dissertations, and peer-reviewed academic journals. The databases I used included Academic Search Complete, Dissertations and Theses at Walden University, Education Research Complete, Education Source, ERIC, Google Scholar, ProQuest EBook Central, and Sage Journals. I employed the following key terms: *long-term English language learners*, *English language learners*, *English language proficiency, academic achievement*, *grades, high school*, *ACCESS for ELLs, WIDA*, *special education*, *content-based language instruction*, *academic language,* and *academic English.* The initial search data criteria included work published within the last 5 years. The initial search produced multiple studies through which I identified other sources after examining reference lists in pertinent articles and books.

Long-Term ELLs

In the United States, 5 million ELLs comprise more than 10% of the student population in public schools (Heritage et al., 2015; Olsen, 2014). Between one-quarter and one-half of all ELLs who enter U.S. schools in primary grades become LTELLs during the course of their schooling experience (Olsen, 2014); thus, most secondary school ELLs are LTELLs. Most LTELLs stall at or below *Intermediate* English language proficiency, as measured by state-adopted ELPAs (Olsen, 2010, 2014). LTELLs are known to have high-functioning social language skills, weaker skills regarding academic

language, and demonstrate significant deficiencies when reading or writing English (Heritage et al., 2015; Olsen, 2014).

Few districts employ focal or formal pedagogical approaches to address LTELLs' linguistic challenges. Often, LTELLs are placed in mainstream courses taught by educators without the tools, skills, or preparation to address the students' specific needs (Olsen, 2014). Because of a lack of remedial support, LTELLs often opt for passivity, invisibility, and nonengagement (Olsen, 2010). Furthermore, many school districts lack the means to identify or monitor the progress of LTELLs (Olsen, 2014; Thompson, 2015). Although a majority of LTELLs want to attend college, they are often unaware that the classes they are permitted or advised to take, their academic skills, and their school records will not allow them to attain that goal (Menken et al., 2012).

Ongoing academic challenges are evinced by the elevated high school dropout and low college-completion rates of LTELLs (Olsen, 2014; Wong-Fillmore & Snow, 2000, as cited in Slama, 2012). Before describing the effects of policies, educators, literacy expectations, and ELPAs on LTELLs' academic progress, it is important to conceptualize this student subgroup. The following section provides insight into the characteristics, definition, identification, and reclassification of LTELLs.

Characteristics of LTELLs

Despite the increasing number of LTELLs in U.S. schools and of federal mandates to address their language barriers, a paucity of research exists on this student subgroup (Menken et al., 2012; Olsen, 2014). Some researchers identified the

demographic characteristics and previous school experiences of these students (Kim & Garcia, 2014; Menken et al., 2012; Olsen, 2010, 2014; Slama, 2012). The following sections detail some of the variables that account for the diversity of LTELLs.

Time in the United States. LTELLs may have been born and raised in the United States or in foreign countries. Those born in the United States often develop conversational fluency in English, but lack academic proficiency (Olsen, 2014). Students may move back and forth between the United States and their countries of origin during their schooling; thus, they are exposed to different school systems and cultures that can affect their bilingualism in a variety of ways (Menken et al., 2012).

Experience with formal schooling. LTELLs may receive inconsistent schooling as they move between bilingual education programs, English as a second language programs, and mainstream classrooms that do not provide formal ELL services (Menken et al., 2012). Other students may remain in ELL programs that do not build upon their home language practices (Menken et al., 2012). Such pedagogical inconsistencies make LTELLs more likely to fail courses and drop out of school than non-LTELLs (Ascenzi-Moreno, Kleyn, & Menken, 2013).

Developmental differences. The rate, pace, and manner with which individual LTELLs develop their language skills and content knowledge vary (Cook & Zhao, 2011; Heritage et al., 2015). LTELLs begin formal schooling at different points along the spectrum of academic language and literacy skill acquisition (Ascenzi-Moreno et al., 2013). Olsen (2014) indicated that LTELLs learn best with teachers who are sensitive and

responsive to them, to the resources they bring to the classroom, and who understand that their knowledge of narrative content, analytical practices, and vocabulary construction develop over time.

Menken et al. (2012) researched the characteristics and prior schooling of LTELLs, emphasizing educational inconsistencies that have contributed to their rapid proliferation in U.S. public schools. The researchers argued that the subtractive quality of students' prior schooling was a major component of their failure to succeed. Subtractive schooling occurs when schools fail to build on linguistic and cultural resources that students already possess. Accordingly, Menken et al. stressed the importance of the adoption of school-wide language polices that provide LTELLs with consistent support to develop solid academic literacy skills in both their first and second languages.

LTELLs in Special Education

The U.S. Department of Education estimated that more than 500,000 ELLs in K–12 schools participate in special education programs (National Clearinghouse for English Language Acquisition and Language Instruction Programs, 2011a, 2011b). Researchers indicated ELL students are at risk of being overrepresented in special education programs because of inappropriate referrals (Cheatham et al., 2014; Sullivan, 2011). For example, Artiles, Rueda, Salazar, Higareda (2005), Sullivan (2011), and Reeves (2006) found that instructors who do not understand ELLs' second language difficulties often mislabel them as *learning disabled*. In addition, teachers who hold negative views of ELLs often refer them for special education services (Cheatham et al., 2014; Harklau, 2000; Sharkey

& Layzer, 2000; Walker, Shafer, & Liams, 2004). Rhodes, Ochoa, Ortiz (2005) and Thompson (2015) found systemically biased procedures and assessment practices increase the chance of ELLs' learning difficulties to be misidentified as cognitive rather than linguistic.

Despite an increase of research on ELLs in special education, few empirical studies pertain to LTELLs in these programs (Kim & Garcia, 2014; Olsen, 2010). Thompson (2015), whose research focused on analyzing the cost and benefits of the LTELL label, presented findings that reflected the language acquisition trajectory, course placement, and learning opportunities of three LTELLs, one of whom also qualified for special education services. The researcher found the academic achievement, the rigor of the course work, and the postsecondary options offered to special education students were far fewer than those open to LTELLs not diagnosed with learning disabilities, and dramatically less than those offered to their English proficient peers.

Brunner (2012), who explored possible ways to improve early identification of LTELLs, provided relevant facts on LTELL study participants classified as special education students. The study included a 2010– 2011 cohort of LTELLs *(n* = 6,394) within the Austin Independent School District (AISD). The AISD identified LTELLs by the number of years they were in school in the United States prior to the Fall term of the current year and their first grade of enrollment in the AISD. Brunner found that of the 2010–2011 LTELL cohort, 84% had entered AISD in early education (grades K–2) and approximately three out of four LTELLs were diagnosed with learning disabilities, with

25% of LTELLS having received special education services, 33% of high school LTELLs being placed in special education, and 34% having repeated one or more grade levels (LTELLs most frequently repeated Grades 1 and 9). Further, only 14% of LTELLs in Grades 6–11 met the criteria to exit the language programs; however, more than 80% of LTELLs who did not exit English language services scored advanced or higher on the speaking and reading portion of the language assessment (Brunner, 2012).

Brunner's (2012) findings provided significant insight as to why a high number of LTELLs do not exit ELL programs within the expected 5- to 7-year period. The researchers validated research on LTELLs and found that, as a group, they tend to be: (a) overrepresented in special education services, (b) misclassified as learning disabled, (c) repeating more than one grade level, and (d) fluent only in the language domains of speaking and reading. The data indicating 80% of LTELLs scored *proficient* on the reading portion of the language assessment was an unexpected finding as it conflicts with existing research that indicates otherwise (Heritage et al., 2015; Olsen, 2014; Wakeman, 2013).

Defining LTELLs

There is no single, formal, working definition of LTELLs in U.S. public schools (Zhao & Maina, 2015). A review of various state school districts' definitions indicated that the LTELL label typically refers to students classified as ELLs for at least 6 or 7 years who have not met state-established English language proficiency criteria (Menken & Kleyn, 2010; Olsen, 2014; Thompson, 2015). Criteria for LTELL classification is often

bound to ELPA results (Gottlieb, 2006). In 2012, the state of California enacted the first definition of LTELLs into law. This new law incorporated the number of years ELLs had been enrolled in U.S. schools and added criteria to evidence a student's struggle with literacy (Thompson, 2015). To be considered an LTELL in California, a student must be enrolled in Grades 6–12 and have resided in the United States for more than 6 years. The student must also have remained at the same level of language proficiency for 2 or more consecutive years, as measured by an English language proficiency test, and have scored *below basic* or *far below basic* on the state content assessment in ELA (Olsen, 2014).

Although California stipulated more than 6 years as the time required to attain English language proficiency, the temporal gamut for the acquisition of English ranges from 5 years in some school districts to 10 years in others (Olsen, 2014). During the course of this study, the WDSD defined LTELLs as ELL students who had been enrolled in U.S. schools for more than 5 years without obtaining English proficiency. District criteria for English proficiency required students to obtain an overall score of 5 or 6 on the ACCESS English language proficiency state assessment and a minimum literacy composite score of 5 (WDSD, 2011).

Some researchers argue that the LTELL label has acquired strong negative connotations, with the description of LTELLs often focusing on students' perceived deficits (Thompson, 2015). Thompson's (2015) examples demonstrated how researchers, whose intent was to draw attention to the educational needs of this ELL subgroup, depicted LTELLs as lacking literacy skills in either language, being disinterested in their

studies, and ready to drop out (Kinsella, n.d.; Menken et al., 2012; Olsen, 2010). Link and Phelan (2013) asserted that any such label entails both costs and benefits and argued that the LTELL label provides access to certain remedial services that may also create stigmas and constrain choices. Thompson (2015) found LTELLs experience stereotyping, loss of status, and discrimination that have a far-reaching effect on their ability to learn. The researcher stressed the importance of ensuring that the construct of language proficiency be accurately assessed and appropriately used to identify, classify, and reclassify LTELLs.

Procedures for Identifying LTELLs

To provide explicit English language instruction to the students who need it, accurate identification and classification of ELLs and LTELLS is essential (Abedi, 2008; Linquanti & Bailey, 2014). Although different states employ different methods of identifying ELL and LTELL students, researchers have identified some procedures that are used nationally (Wolf et al., 2008). Most states administer a home language survey followed by an ELPA (Linquanti & Bailey, 2014). Home language surveys are used to assess the predominant language spoken in a student's home, while ELPAs are employed to determine a student's English language proficiency level with the goal of qualifying him or her for ELL services. All newly enrolled U.S. public school students are required to participate in a home language survey to identify potential ELL students (Bailey & Carroll, 2015). Students whose home language survey results indicate a primary home language other than English are subject to screening tools or placement tests to confirm

their ELL status and determine appropriate levels of English language proficiency and instructional placement.

All but four states rely on some type of home language survey to initially identify students who require further English language screening (Bailey & Kelly, 2013). Bailey and Kelly (2013) recognized that home language surveys are the most common tools used to identify ELLs, although they are not necessarily the most effective tools. According to Linquanti and Bailey (2014), the reliability of home language surveys as a means of accurate identification of potential ELLs is questionable. Missing a screening opportunity may impede the academic progress of students who need services but were not identified as ELLs by the home language surveys (Linquanti & Bailey, 2014). Kim (2011) noted that incorrectly identifying students as ELLs using initial placement scores can lead to their being placed in poor quality ELL or special education programs with detrimental consequences: lower than average academic achievement, higher than average dropout rates, and severely diminished access to higher education.

Some state educational agencies employ the same ELPA they use for annual testing as a screening device to determine which students are eligible for ELL services. Other states use a short placement test for screening. These assessments are subsequent to initial identification by the home language surveys and provide information on students' levels of English proficiency in the language domains of speaking, listening, reading, and writing (Bailey & Carroll, 2015). The purpose of these assessments is twofold: (a) students who fail to meet the English proficiency criteria measured by these assessments

become eligible for ELL services; and (b) students' identification as ELLs sets in motion a trajectory of instruction, also based on their ELPA results.

Procedures and Criteria for LTELL Reclassification

The use of ELPAs is mandated through the ESSA to account for the number of ELL students who have attained the level of English deemed *proficient* and are ready to exit ELL services. ELLs are expected to exit the ELL services after they have achieved academic fluency on state-mandated or state-approved assessments of their English language proficiency. According to research by Cummins (1981b) and others (Cook, Boals, & Lundberg, 2011; Hakuta, Butler, & Witt, 2000; Thomas & Collier, 2002; Zhao & Maina, 2015), the average ELL takes 5–7 years to acquire grade-level academic English language skills. Middle and high school students often take longer to progress from intermediate to full levels of English language proficiency than elementary school students (Cook, Boals, Wilmes, & Santos, 2008; Kieffer, 2008, 2010, 2011). After a student exits an ELL program, he or she becomes a *reclassified*, *re-designated*, or *exited* student. States use these terms interchangeably to indicate ELL students who no longer need ELL support to be fully functional in mainstream English classes.

Each state determines how the ELPA's language domains of listening, speaking, reading, and writing are combined to set their standard. Although effective classification is the goal, in some instances students are reclassified too soon (Slama, 2012) or exited too late (Estrada & Wang, 2013; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Kim, 2011). For example, Slama (2012) found that students who were reclassified during

their early elementary grades experienced academic difficulties later. Estrada and Wang (2013) found students who remained in ELL services too long struggled with remedial coursework and instruction. Similarly, Kim (2011) found protracted time in ELL programs diminished students' academic persistence. Gándara et al. (2003) discovered that cumbersome ELPA procedures associated with reclassification were riddled with inconsistencies and were a major contributor to the current LTELL epidemic. The following section provides an overview of the sociohistorical context of language proficiency based on the second language acquisition theories of Cummins (1979) and Krashen (1982), which comprise the theoretical framework for this study.

Overview of Language Proficiency

Language is at the heart of teaching and learning. It is the medium through which content is accessed and learned, social relationships and identities are formed, and where linguistic competence is constructed (DiCerbo et al., 2014). To be considered language proficient, students must be able to express their linguistic knowledge and language use in the four distinct language domains of speaking, listening, reading, and writing. The domains of reading and listening are considered *receptive* channels that reflect students' abilities to successfully manipulate a wide variety of linguistic input. The domains of writing and speaking are deemed expressive or *output* channels because students use them to communicate ideas and to react to the internal and external stimuli they experience. Students who can effectively use the four language domains to express their

thoughts and derive meaning are considered *language competent* or *language proficient* (Bachman, 1990).

In their study, Cook et al. (2011) established that *English language proficient* entailed three criteria: (a) being proficient on state-required content assessments, (b) being successful with classroom activities, and (c) being able to fully participate in society. The researchers noted that federal law only requires ELLs to be *successful* in academic content rather than requiring them to be *proficient* in academic content. Cook et al. contended that an ELL's *ability* to be successful required her or him to be *proficient* in academic English and content knowledge. The researchers recommended the use of ELPAs (specifically the ACCESS test) to pinpoint the levels of English language proficiency in which most ELLs would be *successful* in passing state content assessments.

Prior to the enactment of NCLB in 2012, state educational agencies did not use ELPA data to determine ELLs' level of English language proficiency. Now, all states use ELPAs to define English language proficiency and to determine if students have achieved full fluency in English. Wolf et al. (2008) examined ELPAs and the levels of language proficiency defined by each state. The researchers found different states defined the levels or stages of language proficiency, through which students were expected to develop English fluency, differently. Differences were found in the numbers of language proficiency levels and the terms used to define each level. The number of levels of proficiency ranged from 3 to 6 (Wolf et al., 2008). The ELPA used in the present study,

ACCESS, described the continuum of language development with five language proficiency levels (1-Entering, 2-Beginning, 3-Developing, 4-Expanding, and 5-Bridging) and a sixth and final exit level called Reaching.

During the timeline of this study, WDSD students were considered English language proficient if they obtained an overall score of 5 or higher on the ACCESS assessment and a minimum literacy composite score of 5. Students at this level of English proficiency are expected to (a) use technical or specialized language appropriate to specific course content areas; (b) produce oral and written sentences that vary in length and linguistic complexity (e.g., stories, essays, or reports); and (c) understand, process, and produce grade-level material that is comparable to that of their English-proficient peers (WIDA, 2013).

Social and Academic Language Proficiency

A defining feature of language proficiency is how language is used to communicate a message. Language can be both social and academic in nature. Cummins (1981a), a well-known scholar in the field of second language acquisition, coined the terms basic interpersonal communication skills (BICS) and cognitive academic language proficiency (CALP) to distinguish between social and academic language. The researcher maintained BICS was easier to acquire because it is cognitively undemanding or *context-reduced*. The use of linguistic cues, such as facial expressions, gestures, objects, and pictures, reduces the complexity and abstraction of BICS. Cummins referred to this as "playground language" (Cummins & Hornberger, 2008, p. 6) because students can

acquire it through daily social interactions of the kind that occur on a school playground. Cummins maintained, however, that CALP is more difficult to acquire because it is complex, context-dependent, cognitively demanding, and more abstract. Reading a science textbook involves using CALP because the only language clues for learners come from the context itself. Students must depend on pictures, charts, and graphs to make sense of the text; hence, the textbook is considered *context-embedded*. If the text also deals with challenging new science concepts, it is *cognitively demanding* (Freeman & Freeman, 2009).

Cummins (1981b) elaborated the BICS and CALP distinction in a quadrants framework that illustrated the differences between BICS and CALP using two linear and intersecting continuums that resembled a cross. The horizontal axis represented the range of contextual support that students need to complete a particular language task or activity, and the vertical axis represented the range of cognitive support needed to complete that same task or activity. The horizontal axis depicted tasks with high context, which students may be familiar with, to abstract concepts that they find more challenging. The vertical axis illustrated tasks that are cognitively undemanding and easier for students, to cognitively demanding tasks that require students to exert their knowledge of CALP (Cummins, 1981b).

Cummins' (1981b) quadrant framework illustrated the type of activities that fit into each of the four quadrants. Cummins defined BICS as oral or written language that occurs when the language is context-embedded and cognitively undemanding (Cummins,

1981b). Activities that are cognitively undemanding and context-reduced were placed in the left upper quadrant and included tasks, such as following directions, interacting in face-to-face conversations, or participating in oral presentations. Activities that are cognitively demanding but are context-reduced were placed in the right upper quadrant (Cummins, 1981b). A telephone conversation, for example, would not be cognitively demanding if the speaker has knowledge of the topic being discussed. However, because the phone conversation is not face-to-face, the physical context would be reduced.

Cummins (1981b) defined CALP as that which occurs when language is context-reduced and cognitively demanding. Activities located in the right lower quadrant required knowledge and use of CALP. To complete these tasks well, a student must rely heavily on her or his academic English proficiency, as knowledge of conversational English would not be sufficient. Cummins emphasized that teachers should ensure activities start in the left upper quadrant and gradually move through each quadrant, ultimately presenting cognitively demanding content in such a way that it is embedded in understandable contexts.

The existing literature on LTELLs appears to support Cummins's (1979) theories by demonstrating that LTELLs often possess well-developed conversational English skills, but lack specialized vocabulary and grammar (Heritage et al., 2015). One of the most pressing challenges for LTELLs is their struggle to understand and respond to assignments in academic texts (Menken et al., 2012; Slama, 2012). This comprehension issue hinders the acquisition of content knowledge in all academic areas (Menken et al.,

2012; Slama, 2012). LTELLs' lack of academic language development and consequent accumulation of academic deficits may be because of the epistemic inconsistency of English language programs to which they are exposed (Abedi, 2007; Olsen, 2010). Factors, such as low socioeconomic status, the absence of positive learning attitudes and motivation, and a lack of formal schooling in their first language, contribute to their failure to acquire adequate grounding in academic English (August & Shanahan, 2006).

Time to Acquire Social and Academic Language Proficiency

Cummins (1979) argued that comprehensible input is what leads students to make sense of the academic language used in content courses, such as social studies, science, mathematics, and English literature. Cummins (1981a) found that students who entered U.S. schools speaking languages other than English received much or all of their early instruction in English. Because the instruction was only partially comprehensible to them, those students failed to acquire an understanding of the academic language or content. When studying immigrant children in Canada, Cummins (1981a) became aware of an interval of several years between students' attainment of peer-appropriate levels of social English skills (e.g., phonology, oral fluency), and grade-level academic norms (e.g., reading and writing skills). Cummins (1981a) also found that when educators conflated students' conversational fluency with their second language (L2) academic proficiency, they actually intensified students' academic difficulties. The BICS and CALP distinction that Cummins' observed in his previous studies were further reinforced in a study by the Toronto Board of Education (Cummins, 1981b), which demonstrated BICS was acquired

within 2 years of exposure to English while CALP required a period of 5–7 years. Research studies conducted in Canada (Klesmer, 1994), Europe (Snow & Hoefnagel-Hohle, 1978), Israel (Shohamy et al., 2002), and the United States (Cook, 2014; Cook et al., 2011; Hakuta et al., 2000; Thomas & Collier, 2002) corroborated these findings, particularly the multiple year disparity between the time required to attain conversational fluency and that needed for grade-level academic language proficiency.

Literacy in First Language Helps Second Language Acquisition

Cummins (2000) contended that students who are effectively instructed in a first language (L1) do better in acquiring a second language. The researcher believed that students can successfully transfer the language skills they develop to a L2 as long as they are motivated to learn the second language. Cummins's claim that concepts are transferable is known as Cummins's interdependence hypothesis. The interdependence hypothesis proposes that common cognitive or academic proficiency is shared by any two languages. Cummins contended that every language has common surface features and underlying those surface features are proficiencies involving more cognitively demanding tasks that are common across languages (e.g., content learning, abstract thinking, problem solving, and literacy). The language used in cognitively demanding tasks involving complex language is CALP, which is transferable across languages.

Cummins (2000) claimed that transfer can occur after a student has developed enough L2 proficiency to comprehend and convey the concepts already learned in the first language. Evidence from considerable research on students in bilingual programs

supports the idea of a common, underlying proficiency. Graves' (2006) review of the literature showed that students who are literate in their first language do better in acquiring English. Freeman and Freeman (2002) also alluded to studies that showed immigrant children with adequate formal schooling in L1 succeeded at higher rates than LTELLs. Freeman and Freeman accounted for the differences in success rates by recognizing that immigrant students with more schooling had already developed academic language in their first language that allowed them to transfer concepts learned from their first language to another.

From theory to practice, the work of Cummins (1981a, 1981b) has significantly influenced the movement towards integrated second language instruction. The researcher's major contribution to this study was the notion that academic language is more cognitively demanding than social language and can take longer than 6 years to acquire. Cummins' differentiation of BICS and CALP created a base for understanding why both the process and the product of content instruction is difficult for LTELLs who develop their social language skills through everyday interactions and activities. Because they are immersed in learning environments where English is the language of instruction, LTELLs acquire social language skills with apparent ease within 6 months to 2 years of daily immersion. Therefore, they can function at high levels in social interactions, but lack the conceptually demanding and cognitively complex language required for learning academic content (Menken et al., 2012; Slama, 2012). Cummins' (1981b) quadrant framework illustrates the distinct differences in the selection and use of words and

thought processes that BICS and CALP activities require of LTELLs. The quadrants address the issue of how to make the oral and written language of content courses more comprehensible. Cummins found that by embedding academic language in comprehensible contexts, environmental clues emerged to make cognitively demanding content easier for ELLs and LTELLs to understand. A later but no less important contribution was Cummins's (2000) interdependence hypothesis, which emphasized that students need to develop their oral language skills in both their native language and in English because skills developed in one language transfer to the other.

Role that Academic Language Plays in Language Acquisition

Understanding and teaching the complex language of academic content that affects LTELLs' academic performance inside the classroom may not be easy, according to different philosophical and methodological perspectives that researchers have regarding academic language. To illustrate the important role that academic language plays in language acquisition and the learning process, researchers must first define the concept of *academic language*. Cummins's (1981b) distinction between BICS and CALP provided one of the first paradigms for academic language acquisition (DiCerbo et al., 2014). Cummins (2000) defined CALP as "the extent to which an individual has command of the oral and written academic registrars of schooling" (p. 67). Cummins's CALP is characterized by specific linguistic features, such as discourse, grammar, and vocabulary, which inform the language domains of listening, speaking, reading, and

writing across a content spectrum that includes language arts, mathematics, science, and history courses (Bailey & Heritage, 2008; Gottlieb et al., 2009; Schleppegrell, 2004).

Other researchers defined academic language differently. For example, Dutro and Moran (2003) described academic language proficiency as an ability to create meaning using oral and written language, connect complex ideas and information, recognize a variety of genres, and use appropriate linguistic strategies to communicate. Díaz-Rico and Weed (2002) considered academic language to be a cognitive set of thinking skills and language abilities used to decode and encode complex concepts. Zwiers (2008) described academic language as discourse strategies, sets of words, and grammar used to describe abstract concepts, higher order thinking, and complex ideas. Gottlieb and Ernst-Slavit (2013) viewed academic language as a developmental construct for all students that increases in complexity and sophistication in a vertical chronology. For LTELLs, developing academic language has a horizontal dimension that spans from one level of language proficiency to the next (Gottlieb & Ernst-Slavit, 2013).

Bailey and Heritage (2008) fine-tuned the concept of academic English by breaking academic language into school navigational language and curriculum content language. School navigational language describes the language used by students to communicate with teachers and peers in the school setting and curriculum content language describes the language used in the process of teaching and learning content material (Bailey & Heritage, 2008). Scarcella (2003) expanded Bailey and Heritage's description of school navigational language, adding the notions of foundational

knowledge of English and essential academic language. The researcher described foundational knowledge of English as the basic skills and common vocabulary required for communication in both academic and nonacademic settings, whereas knowledge of essential academic language requires explicit, systematic instruction that cannot be compared with the social, informal language, which Cummins referred to as BICS. Scarcella described essential academic language as the academic language used across all content areas, including discourse features, complex sentence structures, and academic vocabulary. The researcher posited that all three types of English knowledge were necessary for students to acquire curriculum content language and argued that ELLs should already possess measurable cognitive and language skills, such as mastery of foundational knowledge of English, demonstrable knowledge of school navigational language, and expanding control of essential academic language, prior to the study of course-specific academic English (Scarcella, 2003).

Academic Language and the Common Core State Standards

The Common Core State Standards for ELA have also shaped ideas of academic language (Zwiers, 2014). These standards are college and career-ready standards aimed at ensuring students graduate from high school well prepared to succeed in the interdependent domains of higher education and professional competence (Heritage et al., 2015). With regard to language functions and thinking skills, the Common Core State Standards are concerned with students' abilities to support ideas with textual evidence, determine how themes and main ideas are conveyed through particular details in the text,

determine cause and effect, and finally, analyze how chapters, sentences, and words contribute to the development of the key ideas in texts (Zwiers, 2014). These standards exemplify the variety of ways that students are expected to use academic language to assimilate knowledge and to demonstrate learning in ELA. The academic language referred to throughout the Common Core State Standards serves three interrelated and wide-ranging functions, including language complexity, higher-order thinking, and abstraction (Zwiers, 2014). These functions overlap and take on different forms depending on the grade level and the content area in which they are taught (Zwiers, 2014).

Language complexity. The first aspect of academic language is complexity (Zwiers, 2014). Schleppegrell (2004) noted that one of the main functions of academic language is to describe complex concepts as clearly as possible. For example, in language arts, complex relationships exist between characters, plots, and literacy devices, and these complex relationships must be organized and expressed in writing. Complexity deepens as students ascend to higher levels in their education and careers. In the professional world, employees must cultivate their relationships with coworkers, give presentations, create projects, conduct lab experiments, read financial reports, etc. Professionals must strive to use the tools of language to make communication for work-related tasks clear and accessible to colleagues and supervisors (Zwiers, 2014).

Higher-order thinking. Academic language is also used to foster higher-order thinking (Zwiers, 2014). Students use academic language to describe complex

epistemological processes, often called *higher-order thinking skills*. These processes are used to express ideas, comprehend information, and solve problems (Facione, 1990; Swartz, 2001). Many lists of thinking skills have emerged, led by Bloom, Engelhar, Furst, Hill, and Krathwohl's (1956) famous taxonomy. Bloom et al. proposed six vertical levels of thinking, starting with knowledge and progressing to subsequent levels of comprehension, application, analysis, synthesis, and evaluation. Other researchers argued that academic language is used for other cognitive functions, including applying, analyzing, classifying, comparing, explaining, evaluating, hypothesizing, informing, justifying, and predicting (Valdez-Pierce & O'Malley, 1992; Wiggins & McTighe, 1998). A cursory analysis of the Common Core State Standards English language development standards (Zwiers, 2014) revealed a wide range of thinking skills that students must possess. For example, the ELA and Literacy standards of the Common Core State Standards emphasize building on the ideas of others, interpreting themes, evaluating evidence, supporting arguments, and synthesizing ideas from multiple sources (National Governors Association Center for Best Practices, Council of Chief State School Offices, 2010).

Abstraction. The third function of academic language is to describe abstract concepts, which are ideas or relationships that individuals cannot easily illustrate with images, point to, or act out, as is often done in school, home, and social settings (Zwiers, 2014). To illustrate an abstract concept, Zwiers (2014) used the sentence, On the other hand, the two scientists had differing views on the topic of evolution (p. 28). The term,

views, refers to the scientists' abstract thoughts on evolution as opposed to perspectival vistas. Students, especially ELLs and LTELLs, may require an explanation of such homonyms. They must also be taught that the prepositional phrase *on the other hand* often announces a textual detour as do the conjunctions *but*, *however*, and *yet*. This abstract compare-and-contrast thinking, which may seem automatic for L1 readers accustomed to such syntax in their native idiom, is challenging for L2 learners, no matter how bright they are. One reason ELLs and LTELLs find abstract concepts difficult is that these concepts often defy visual representation (Snow & Brinton, 1997).

Although the Common Core State Standards outline the expectations of what students should be able to understand and do within each school discipline, WIDA researchers and developers are concerned about LTELLs understanding the academic language used in schools and accelerating their ability to make meaning from that language. The WIDA standards' framework includes English language development standards that inform the ACCESS assessment, describe the academic language used within the classroom, and delineate a clear path for students to acquire the academic language needed to master grade-level content (Gottlieb, 2013).

Second Language Acquisition Process

The variety of ways that students are expected to use academic language to satisfy Common Core State Standards and WIDA standards requires them to be proficient in the English language, not only at the interpersonal level (BICS), but also at the academic level (CALP; Cummins, 1979, 1981a; Cummins & Hornberger, 2008). This is a cause for

concern among stakeholders dealing with LTELLs, especially in school districts where teachers often lack the pedagogical methods and practices that have proven beneficial to LTELLs' acquiring academic proficiency. The WDSD, which adopted the Common Core State Standards in 2014 and the WIDA ACCESS assessment in 2012, is such a district. The supply of teachers trained to work with LTELLs is inadequate. Less than 5% of the district's teachers have Teaching English as a Second Language endorsement, yet 98% of them have ELLs in their classrooms (WDSD, 2016).

Teachers need to have some theoretical and methodological knowledge about second language acquisition to become effective instructors of pupils in the act of literacy development. Researchers indicated such awareness will support teachers in developing their students' academic language proficiency well enough to meet the intellectually demanding goals of the new standards and language assessments (Heritage et al., 2015). Cummins (1979) and Krashen (1982) are well-known scholars in the field of second language acquisition who have developed theories applicable to the pedagogy of ELLs, a category that includes LTELLs. The following section introduces Krashen's (1982, 1985) monitor model, which offers insight into the psychological aspects of second language acquisition. The five hypotheses of Krashen's monitor model have significant relevance when theorizing about how to help LTELLs acquire content-rich academic language.

Krashen's Monitor Model

Behaviorist B. F. Skinner provided one of the earliest explanations for language acquisition, claiming that children learn language based on stimulus-response-

reinforcement principles in which they receive positive reinforcement for correctly associating words with meaning (Ambridge & Lieven, 2011). Skinner's perspective was heavily critiqued by theorists like Chomsky (1959), who argued that children do not merely learn language by imitating those around them. Chomsky theorized that humans are born with a language acquisition device that provides them with a natural ability to process linguistic rules. The researcher contended that a language acquisition device is an innate faculty that guides children in acquiring language while routinely generating novel sentences and rules without instructional input (Mitchell, Myles, & Marsden, 2013).

Although the work of both Skinner and Chomsky focused on primary language attainment, it triggered significant changes in the ways scientists and scholars viewed second language acquisition. Chomsky's belief in the innate linguistic aptitude of young children served as a framework for Krashen's (1982) model of second language acquisition, the monitor model. Krashen's monitor model is important to this study because it is used to explain the physiological and cognitive processes LTELLs traverse while learning a second language. The five hypotheses contained in Krashen's model are the acquisition-learning, natural order, monitor, input, and affective filter hypotheses.

Krashen (1985) regarded the acquisition-learning hypothesis as the most basic of the five hypotheses. The researcher believed two independent systems of L2 performance exist, one acquired and the other learned. The acquired system refers to the subconscious process of internalizing a second language naturally, such as the processes through which children acquire their first language (L1). Krashen postulated that this type of second

language learning is accumulated subconsciously. Alternatively, the learned system refers to the conscious process of language learning, such as learning how to apply grammatical and syntactical rules.

Krashen (1985) noted that L2 learners rely heavily on the learned system at school, although it is limited and devoid of the naturalistic interactions required for the subconscious development of a second language. The researcher believed that for most learners, the challenge of L2 acquisition is moving from the learned system to the acquired system, which requires extensive exposure to and use of L2 in quotidian and academic circumstances. The classroom implications of Krashen's learning-acquisition hypothesis require that LTELL instructors foster language acquisition by creating opportunities for students to work together and engage in natural and meaningful social interactions (Latifi, Ketabi, & Mohammadi, 2013). Swain (1995) noted that L2 learners process language in more profound and lasting ways if they are allowed to frequently and freely engage in L2 classroom conversations with their classmates.

According to Krashen's (1985) natural order hypothesis, most learners acquire the grammatical structures of language in a predictable order. Krashen reported that learners acquire the progressive forms, plural forms, and linking verbs first, followed by the articles and the progressive auxiliary. Learners then acquire the irregular past tense and finally the possessive, regular past tense, and third person singular. Studies that reinforced the existence of a *natural order* of language acquisition indicate that the order in which the rules of language are acquired is independent of a learner's age, first

language background, and conditions of exposure (Dulay & Burt, 1974; Fathman, 1975; Makino, 1980, as cited in Krashen, 1982). The implications of the natural order hypothesis are that LTELLs will not necessarily acquire English in the order it is taught, but in the natural way that the brain learns language. Hence, it is important for teachers to understand that LTELLs will acquire English more rapidly when they are engaged in language-rich, supportive, and culturally respectful classroom experiences than through explicit grammar instruction (Akhavan, 2006; Hoover & Patton, 2005).

In the third hypothesis, the monitor hypothesis, Krashen (1982) specified differences in the ways acquisition and learning are used in L2 performance. The researcher hypothesized that students are able to use their previous understanding of language rules to self-correct and monitor their language output. According to the third hypothesis, the student learns to monitor or edit utterances either before or after writing or speaking them. Krashen appealed to the concepts of the monitor hypothesis to explain individual differences in the way learners use the target language to communicate their purpose. The researcher suggested it was possible to find people who monitored too much, too little, or optimally. People who monitor too much may sacrifice speed and fluency for content accuracy, while those who monitor too little may sacrifice content accuracy for speed and fluency. Optimal monitor users, on the other hand, understand how to use monitors to communicate effectively without sacrificing content accuracy, speed, or fluency. Understanding that language processing and production is largely a

subconscious activity led other researchers to explore interactions between implicit and explicit learning (Mitchell et al., 2013).

With Krashen's (1982) fourth *input* hypothesis, the researcher attempted to explain how a language learner internalizes knowledge. According to this hypothesis, language is only acquired when one is exposed to written or spoken input that is comprehensible. Like Cummins (1979), Krashen stated that for optimal language acquisition, the language used must be both comprehensible and somewhat more advanced than a student's current level (i+1, where i stands for language input and $+1$ represents the language level the student is striving to reach). Input that is too simple (material with language that has already been acquired) or too complex (material too difficult for intuitive comprehension) is not useful for language acquisition. Krashen recognized that not all learners could be at the same level of linguistic competence at the same time; hence, the researcher suggested designing syllabi that ensure all learners received some $i + 1$ for their current stages of language competence. A key claim of Krashen's input hypothesis is that learners do not need to speak to develop language. Rather, regular opportunities to parse and interpret suitable $i + 1$ input, in which learners simply absorb a second language, are sufficient for L2 acquisition (Krashen, 1982).

In the final or *affective filter* hypothesis, Krashen (1982) hypothesized that speakers' anxiety, self-esteem, and lack of motivation impedes the process of absorbing *input* and transforming it to *intake*, thus creating barriers to L2 acquisition. Krashen believed these affective filters limit the level and quality of speakers' receptiveness.

Students who suffer from high anxiety and low self-confidence seek less input and develop strong affective filters that prevent input from reaching the part of the brain responsible for language acquisition. In contrast, students who implicitly trust in their L2 abilities remain open to all types of linguistic influx, agglomerating and refining the information they acquire (Krashen, 1982). Some scholars criticized Krashen's (1982) affective filter hypothesis as vague and atheoretical (Brown, 2007; Zafar, 2009). However, recent brain researchers have substantiated Krashen's hypothesis, finding that anxiety, low self-esteem, and motivation can affect working memory and prevent it from retrieving stored information (Sousa, 2008).

Critical Reactions to Krashen's Hypothesis

Researchers have leveled a considerable amount of criticisms against Krashen's five hypotheses (1983). For example, McLaughlin (1990) refuted Krashen's belief that learning and acquisition were not bidirectional, arguing that learned knowledge (from instruction) can lead to improvements in the process of L2 acquisition (Norris & Ortega, 2000). Zafar (2009) stated Krashen's natural order hypothesis failed to account for the considerable influence of students' L1 fluency on their acquisition of a second language. Zafar pointed to study results that indicated confident, complex self-expression in the mother tongue trumps learning grammatical structures in predictable sequences. Researchers (Ellis, 2003; Rubin, 1975) also criticized Krashen's monitor hypothesis for relegating language monitoring to a post-learning process (Latifi et al., 2013). Rubin

(1975) claimed monitoring was a basic strategy of L2 acquisition, and Ellis (2003) described monitoring as fundamental to successful language learning.

Krashen's (1982) input hypothesis has also undergone considerable scrutiny by other scholars. For example, Swain (1995) countered Krashen's claim that output was not necessary for language acquisition, arguing that just taking in a language was not enough to acquire it. According to Swain, L2 learners need many opportunities to use a target language. The researcher contended that content lessons should elicit long and complex utterances whereby learners practice grammatical and semantic coherence, as opposed to short responses of minimal complexity (Allen, Swain, Harley, & Cummins, 1990; Ellis, 2003). Similarly, Hernandez (2003) believed output was necessary to give L2 learners opportunities to negotiate meaning through the use of verbal and nonverbal strategies. Long (1983) also viewed the negotiation of meaning as a critical aspect of language acquisition.

Moreover, the novelty of Krashen's $i + 1$ formula was faulted for being too close to what Ausubel (1963) proposed as *meaningfulness* or *subsumability* (Brown, 2007). Ausubel emphasized the importance of engaging students in meaningful tasks, postulating that teachers could enhance the meaningfulness of new material by associating it with students' existing knowledge. From Ausubel's perspective, activating relevant schema or conceptual patterns prepares a student's cognitive structure for learning and allows new material to be subsumed or incorporated into her or his cognitive structure. Ellis (2003) also challenged the notion that L2 acquisition could occur through

absorption of *i* + 1 alone, arguing input is necessary but not sufficient for acquisition of language to take place.

A final critique was levied against Krashen's *affective filter* hypothesis. Zafar (2009) disputed Krashen's claim that children do not possess the affective filters that prevent most adult L2 learners from completely acquiring a second language. The researcher argued that children also experience variability in anxiety, motivation, and self-confidence that may prevent them from naturally acquiring a second language. Brown (2007) alluded to literature demonstrating many cases in which adults were able to acquire quasi-native proficiency, thus providing evidence that Krashen's hypotheses may be flawed.

Relevance of Krashen's Work

A considerable amount of criticisms surrounds Krashen's (1982) hypotheses. Still, the monitor model and five stages of language acquisition have significant relevance when theorizing optimal instructional practice for LTELLs. Krashen (1982) maintained that adolescents have the ability to acquire a second language in the same natural way that they acquired their first language. Hence, the proper way to teach them an additional language is by providing plenty of input in a safe and dialogue-rich learning environment. Students exposed to sufficient comprehensible linguistic messages may acquire a second language relatively effortlessly. The critical element, according to Krashen, is for teachers to provide the right amount of input. They can do this effectively by giving students plenty of contextual clues through body language, physical objects,

and pictures (Krashen & Terrell, 1983). Additionally, Krashen's *i* + 1 concept suggests that teachers who make use of scaffolding strategies that embed language in context can propel students to advance to the next level of achievement. Teachers who understand how to create noncompetitive environments that are encouraging and rich in visual and auditory stimuli can lower students' affective filters and provide LTELLs with diverse opportunities to experience academic success. The more success students experience, the higher their self-confidence. The higher their self-confidence, the more motivated students become to master a second language.

Bridging English Language Proficiency to Academic Achievement

Krashen defined an optimal learning environment for LTELLs as being one that provides them the opportunity to receive academic language in a meaningful context. Alford and Niño (2011) suggested that providing LTELLs with multiple opportunities to receive and produce academic language is the two-way practice essential for language development and for deepening the meaning of academic course content. The focus on their achieving rigorous standards makes it essential that LTELLs' language acquisition corresponds with their academic achievement and that they learn the concepts and skills of the content areas simultaneously.

Gottlieb, co-founder and lead developer of WIDA, published two books on ELL assessments, *Assessing English Language Learners: Bridges From Language Proficiency To Academic Achievement* (2006) and *Assessing English Language Learners: Bridges To Educational Equity* (2016), both of which are meant to enable educators to understand

that L2 learners' language proficiency is an inextricable component of their academic ability. To bridge the gap from language proficiency to academic achievement, Gottlieb recommended that educators pair language objectives and content objectives so that students recognize both as integral parts of learning and using academic English. Given that the English language development standards that anchor the ACCESS assessment align with the academic content standards adopted by the WDSD and taught by district English teachers, LTELLs' ELPA scores should predict their academic achievement grades in courses conducted in English.

What English Course Grades Measure

The content knowledge of LTELLs in English courses is assessed via academic grades. In most public schools, academic content standards are the exclusive source and anchor for measuring students' academic achievement. Thus, the primary objective for students in English classrooms is to gain knowledge about curricular content and meet the academic content standards. For LTELLs, acquiring the language of content (academic language achievement) must correspond with learning the concepts and skills of the content area (academic achievement). As students advance from grade to grade, they master successive ELA standards in speaking, listening, reading, and writing and begin to exhibit the ability to demonstrate independence, build strong content knowledge, and respond to varying demands of audience, purpose, task, and discipline. Moreover, students acquire the capacity to comprehend as well as critique, value evidence, use

technology and digital media strategically, and understand new perspectives and cultures (WDSD, 2016).

To measure achievement, teachers must consistently assess students' progress. Although there are different means of communicating what students learn, issuing a letter grade for each academic subject is a standard procedure in most public middle and high schools (Allen, 2005). In the WDSD, achievement grades in high school English courses are expressed as letter grades. The WDSD computes grade point averages (GPAs) and class ranks based on the symbols and scales depicted in Table 1 (WDSD, 2006). The grade a student receives is meant to represent the teacher's best judgment of how the student performed relative to the explicit learning objectives of the English course.

Table 1

WDSD Secondary School Grading Scale (Grades 6–12)

Letter Grade	Performance	%	GPA Weight
A	Excellent	90–100	4.0
B	Above Average	80–89	3.0
C	Average	70–79	2.0
D	Below Average	60–69	1.0
F	Failure	Below 60	0

The major reason for assigning grades is to create a public record of students' academic achievement that can be used by students, parents, teachers, guidance counselors, school officials, post-secondary educational institutions, and employers to gauge levels of content mastery (Airasian, 2000; Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000; Nitko, 2001; Oosterhof, 2001; Stiggins, 2001). Allen (2005) confirmed

the assignment of grades can expand learning opportunities. Students who obtain high letter grades gain admission to universities and colleges and receive tuition assistance and scholarships (Allen, 2005). Pintrich and Schunk (2002) demonstrated that when students received grades that accurately depicted their true levels of academic knowledge, they believed in their abilities to succeed academically, demonstrated better self-efficacy, and were more motivated to learn. These positive attributes, which lower the affective filter, are optimal for second language acquisition (Krashen, 1982).

Concerns of whether student grades accurately and effectively communicate meaningful information about students' academic achievement have been raised for more than a century. Efforts made to help teachers understand the effective function and purpose of grades for evaluating students appear throughout the academic literature (Airasian, 2000; Brookhart, 1993; Cross & Frary, 1996; Gredler, 1999; Linn & Gronlund, 2000; Marzano, 2000; O'Conner, 1995; Stiggins, 2001). Allen (2005) recommended that educators experience opportunities to collaborate, create, and implement evaluation and grading plans that are valid. The researcher argued that providing assessments and grading workshops helps practicing teachers develop appropriate criteria to assign grades that reflect students' actual academic achievement.

To ensure that high school English teachers provide grades that are accurate indicators of student achievement, the WDSD implemented professional learning communities. The WDSD requires educators to be familiar with state-adopted ELA content and English language proficiency standards for the courses they teach, and also

be familiar with students' academic, behavioral, and English language proficiency data. The WDSD English teachers are also required to participate in professional learning communities for which they are paid up to 6 hours a month to develop common homework assignments, assessments, and grading practices for their content areas. This practice allows educators time to collaborate and work alongside their school administrators to determine course content and establish guidelines for the assessment of English content learning. Research demonstrates a strong correlation between teachers establishing and working toward rigorous, targeted goals and student achievement (DuFour, DuFour, Eaker, Many, & Mattos, 2016). Professional learning communities provide a framework for identifying and prioritizing needs, instructional planning, progress monitoring, and consistent grading that positively influences students' language development and academic achievement in English courses.

Language Implications of Letter Grades

Researchers have not explored the predictive relationship that exists at any grade level between LTELLs' English language proficiency and the letter grades they receive from their teachers in ELA classes. A number of researchers at international secondary schools, colleges, and universities have explored the relationship between English language proficiency and students' academic performance. Findings from their research indicate that poor English language proficiency can drastically hinder ELLs' post-secondary GPAs. For example, Ghenghesh (2015) investigated the relationship between English language proficiency and academic performance among 566 English as a foreign

language students enrolled in engineering, business, and computer science classes at a British University in Egypt. Ghenghesh used data from the International English Language Testing System and students' GPAs. Colleges and universities worldwide use the English Language Testing System to test and gauge students' English language proficiency in speaking, listening, reading, and writing. Study results indicated a strong relationship between English proficiency and participants' academic success. Specifically, the researcher found that students with higher English proficiency scores performed better in their degree area courses while those with lower proficiency had the highest rates of failure. Ghenghesh's study substantiated the findings of previous scholars (Maleki & Zangani, 2007; Roche & Harrington, 2013; Sadeghi, Kashanian, Maleki, & Haghdoost, 2013; Sivaraman, Balushib, & Rao, 2014).

In addition to research on ELLs, many researchers have explored the relationship between English language proficiency and the academic performance of English as a second language learners. For example, Sahragard and Baharloo (2009) conducted a correlational study of students majoring in English language and literature at Shiraz University in Iran. The researchers found that students who demonstrated higher levels of English language proficiency performed better in their classes, as indicated by higher GPAs. Sadeghi et al. (2013) also found that English proficiency significantly affected the academic achievement of medical students.

In another study conducted at an Australian University, researchers who investigated the relationship between English language proficiency and academic

achievement, as measured by participants' GPAs, found a significantly positive relationship between English language proficiency and GPAs (Feast, 2002). Similar results were found by researchers in Melbourne, who also identified a strong correlation between students' GPAs and English Language Testing System scores (Hill, Storch, & Lynch, 1999). Researchers in higher education institutions in the United States, Nigeria, Malaysia, and India have also found a direct relationship between international students' English language proficiency and their academic performance (Aina et al., 2013; AlHaddad, Mohamed, & AlHabshi, 2004; Kumar, 2014; Xu, 1991).

Few researchers have focused on investigating the relationship between English language proficiency and academic performance of students at the secondary school level. Wilson and Komba (2012) conducted a study in a secondary school in Tanzania. The results of the study revealed that English proficiency had a significant effect on the academic performance of the participants. This finding is in agreement with Fakeye (2014), who studied Nigerian English as a foreign language students and revealed a significant correlation between the English language proficiency and academic achievement of the subjects studied.

English Language Proficiency Assessments

Prior to the enactment of NCLB, public school districts were not federally mandated to monitor the consistency or effectiveness of their schools in meeting the educational needs of ELLs (Bailey & Carroll, 2015). The reauthorization of the Elementary and Secondary Act, also known as the NCLB (2001), introduced new

expectations for the development of academic English language proficiency among ELLs. The changes in federal policy motivated educators and state education agency researchers to explore the meaning of English language development in U.S. schools, as well as the manner in which English language proficiency was conceptualized and measured (Forte et al., 2012). These changes also emphasized measuring the kind of language proficiency necessary for academic success (Bailey & Wolf, 2012).

Title I of NCLB (2001) mandated state education agencies to administer assessments that evaluated the English language proficiency of ELLs. Title III of NCLB required that:

1. Annual assessments align with English language proficiency standards that support ELLs' access to academic content standards. The English language proficiency standards should be linked to states' academic content standards (Bailey & Wolf, 2012).
2. Annual assessments yield separate language proficiency scores in speaking, listening, reading, writing, and comprehension. The comprehension score is a composite of the reading and listening scores (Bailey & Carroll, 2015).
3. Language proficiency scores evaluate the progress that ELLs make towards acquiring English and the degree to which they have achieved proficiency in each of the four assessment categories: speaking, listening, reading, and writing (Forte & Faulkner-Bond, 2010).

The NCLB Act challenged educational leaders in each state to develop an integrated system of objectives, ELPAs, and English language proficiency standards linked to academic content and student achievement standards (National Research Council, 2011).

ACCESS Assessment

The WIDA Consortium was created in response to the NCLB requirements pertaining to standards and assessments for ELLs. A fundamental objective of the WIDA Consortium was to create accountability measures for academic language proficiency in accordance with federal and state policies (WIDA, 2015). Originally established in 2003 with funding from a U.S. Department of Education enhanced assessment grant, the WIDA Consortium was formed to (a) develop a standards and assessment system, (b) plan support for continuing English language development, and (c) satisfy legal assessment and accountability requirements described in the NCLB Act (Fox & Fairbairn, 2011). The acronym for WIDA evolved from the three states that were initially involved in the grant: Wisconsin (WI), Delaware (D), and Arkansas (A). When Arkansas dropped out at the last minute, the phrase *World-Class Instructional Design and Assessment* was created to coincide with the existing acronym (WIDA, 2014b).

The WIDA English language development standards align with state academic content standards that form the core of the WIDA Consortium's approach to instructing and testing ELLs. In 2005, WIDA developed a diagnostic tool called Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs). In accordance with NCLB and ESSA policy, ACCESS is

a standards-based English language proficiency test designed to measure the social and academic progress of ELLs. Because ACCESS for ELLs was the focus of this study, it is important to understand exactly what the assessment tests measure and how they are scored. ACCESS is used to measure ELLs' social and instructional aptitude when dealing with vocabulary required in the content areas of language arts, science, mathematics, and social studies classes across the four language domains of speaking, listening, reading, and writing (WIDA, 2014a). As of 2015, 37 state education agencies adopted and administered ACCESS to more than 1,372,611 ELL students in K–12 classrooms throughout the United States (Fox & Fairbairn, 2011; WIDA, 2015).

The speaking section of the ACCESS for ELLs is adaptive and assessed through scripted, face-to-face interactions. Local school agency test administrators score the assessment using grade-appropriate speaking rubrics. Students' listening and reading skills are assessed through multiple-choice questions. On the writing section, students receive three or four writing tasks, depending on which test tier they are on. Apart from the kindergarten exam, the writing portion of the ACCESS for ELLs is centrally scored by MetriTech's trained raters, who utilize a 6-point writing rubric to determine the total raw scores (WIDA, 2015). MetriTech combines the raw scores for the language domains of speaking, listening, reading, and writing for all grades and converts the scores to scale scores and proficiency levels used for school reporting (WIDA, 2015).

Reporting ACCESS Scores

ACCESS results are reported in three ways: as raw scores, scaled scores, and English language proficiency scores. Raw scores indicate the number of correctly completed items or tasks out of the total number of items or tasks. Scaled scores report on grade levels in relation to the continuum of language development, which allows state educational agencies to compare student progress across grade levels. English language proficiency scores are used to interpret scaled scores in terms of language proficiency levels that range from 1 (*entering*) to 6 (*reaching*; WIDA, 2015).

The overall ACCESS scale score, or composite score, is weighted based on students' scores in speaking (15%), listening (15%), reading (35%), and writing (35%) (Fox & Fairbairn, 2011; WIDA, 2014a). The weighting scheme is supported by WIDA's (2015) guiding principles of language development and the research-based perspective that students develop language proficiency in these four areas interdependently, at different rates, and in different ways (Gottlieb & Hamayan, 2007; Spolsky, 1989; Vygotsky, 1962). Reading and writing are weighted more heavily because they are considered part of the academic domain, whereas listening and speaking skills theoretically start before the child is enrolled in school and are strengthened by daily use, in and out of class. Students' scaled scores are calculated and reported for each of the four language domains and four composite scores. The lowest possible scaled score is 100 and the highest is 600. Scaling the scores makes it possible to compare test scores within any of the four domains, across grades and tiers, on a single vertical scale.

Because separate scales exist for each domain, scaled scores cannot be compared across domains. For example, a scaled score of 400 in listening is not the same as a scaled score of 400 in reading (WIDA, 2014a, 2015).

ACCESS Uses

Since the enactment of NCLB, state educational agencies and test designers have developed, validated, and refined ELPAs to satisfy legal assessment and accountability requirements described in the NCLB Act (Bailey & Carroll, 2015; Fox & Fairbairn, 2011). Assessment results can be used to produce or enhance the desired outcomes of successful achievement for ELL students. In addition to identifying and reclassifying ELL students, the following section reviews four identifiable ELPA uses, including (a) confirming ELL status, eligibility for ELL services, and initial program placement; (b) measuring ELLs' progress towards English language proficiency; (c) monitoring annual English language progress to guide instructional practices; and (d) supporting special needs students' individualized education programs.

Confirming ELL status, eligibility for ELL services, and initial program placement. Some state education agencies use results from ELPA tests to determine students' eligibility for ELL services. Other agencies use a short placement test or screener. These assessments follow students' initial identification as ELLs through the home language survey and provide information on their levels of English proficiency (Bailey & Carroll, 2015). The information from these assessments determines a twofold educational purpose. Students who fail to meet the English proficiency standards set by

these assessments become eligible for ELL services. This placement also sets in motion the trajectory of instruction provided to students.

Measuring ELLs progress toward English language proficiency. Prior to the enactment of the ESSA law (2017), Title III of the NCLB of 2001 required state education agencies to develop progress attainment benchmarks for ELLs, called Annual Measurable Achievement Objectives (AMAOs). The AMAOs were based on annual assessments of ELLs' English proficiency in the language domains of listening, speaking, reading, and writing. State educational agencies had the discretion of defining and setting targets for AMAO progress. From 2012 until 2016, ACCESS was used in WIDA Consortium states to determine whether state school districts met their annual AMAO targets. Seven states that did not belong to the WIDA Consortium either developed their own ELPAs, modified an existing assessment, or used a commercially available ELPA to measure their AMAO targets. Under the NCLB Act, all Title III funded educational agencies were obligated to meet AMAO targets annually. Sanctions were levied against educational agencies that failed to meet AMAO targets in any given year.

The recently enacted ESSA law (2017) eliminated the AMAOs that tracked ELL outcomes during the NCLB era. Instead, ESSA requires all states to use ELPAs to measure increases in the percentage of ELLs who are making progress in achieving English language proficiency (Carnock, 2016). Educational leaders from each state may choose which ELPAs they use to measure the required progress and have discretion to define progress and set achievement targets for those ELPAs (Carnock, 2016). Public

school districts belonging to the WIDA Consortium, including the WDSD, will continue to use ACCESS to measure ELL progress toward English language proficiency as set forth by ESSA law.

Monitoring English language progress to guide instruction. Educators use ELPA results to monitor the progress of their students' English language literacy development. To improve students' content knowledge and skills, Bailey and Carroll (2015) recommended that teachers apply an approach that identifies the key practices and disciplinary core ideas in the new content standards and the receptive and productive language functions that students need to carry out those practices. Their approach is similar to Gottlieb's (2016) recommendation to pair language and content objectives to simultaneously develop students' language skills and content knowledge.

Supporting individualized education programs. State educational agencies use ELPA results to determine students' eligibility for special education services and to write developmental goals that become part of their individualized education program. The programs describe the types of services and supports provided to students eligible for special education services. Students' special education identifications and educational services are usually based on data gathered and analyzed by educators and other professionals (Liu, McGhee, & Kushner, 2011). English language proficiency scores are part of the data used to make decisions. Educators who do not understand the time-consuming complexity of second language acquisition and how it can affect L2 students' academic progress are more likely to inappropriately label ELLs as learning disabled

(Reeves, 2006).

Cheatham et al. (2014) used ELPA scores to explore whether teachers' knowledge of students' English language proficiency affected their decision to refer them to special education services. The study participants included 214 preservice teachers. Study participants were asked to rank seven possible causes of students' classroom difficulties. The information provided to the teachers varied depending on which of three groups they were assigned (Cheatham et al., 2014). The results of the study indicated that participants who received the most information on students' English language proficiency levels were more likely to correctly attribute classroom difficulties to limited English proficiency. Even then, 35% of the preservice teachers attributed students' learning difficulties to their motivational, intellectual, behavioral, or emotional shortcomings (Cheatham et al., 2014).

Ability of ACCESS to Predict Academic Achievement

The use of assessment data to inform instructional decisions is not new. Joan Cele, a famous teacher and director of a Latin school in Zwolle (a city of the province of Overijssel, Netherlands) first introduced examinations into the school curriculum in the 14th century and used assessment results to determine student grade-level promotions (Wilbrink, 1997). Cele was credited with establishing the European model of the graded school, which uses assessments to promote students and ranks students based on merit (Wilbrink, 1997). In the 21st century, assessment data are still used as a predictor of students' future achievement. Some researchers study the use of assessment data to identify students for remediation or intervention programs (Heatley, 2002). Other

researchers explored the predictive value of assessments to earmark students for interventions or determine specific instructional strategies on high school exit exams, end-of-course exams, or norm-referenced tests (Georgia Department of Education, 2008; Gewertz, 2007; Parker et al., 2009; Wakeman, 2013). In this study, I tested the hypothetical proposition that ACCESS predicts academic success, as measured by letter grades. As previously mentioned, research on LTELLs is limited. Although research is lacking on the predictability of ACCESS to determine the developmental progress of LTELLs in high school English courses, or any other secondary course content, the following research demonstrates that the predictive value of ACCESS has been noted in research pertaining to ELLs.

Wakeman (2013) analyzed test scores of 164 ninth-grade ELL students from one high school in Georgia to determine if a predictive relationship existed between ACCESS test scores and the End-of-Course Test (EOCT) in biology. Latino students constituted 94% of the sampled population ($n = 154$). Only one ELL student in the sample had a designated disability and no LTELLs were mentioned. Wakeman found a strong correlation between the overall subscale score of the ACCESS and the EOCT scores in biology, $r = .414$, $n = 164$, $p = .001$, showing ELLs with higher English language proficiency scores received high scores on the EOCT in biology. The researcher also found that the four language domains of speaking, listening, reading, and writing had a positive correlation with EOCT scores in biology, with the writing subscale score having the highest predictive value of the four.

The speaking domain contributed the least to the EOCT scores in biology. The ACCESS literacy subscale scores, which encompassed reading and writing, proved to be the most predictive of ELLs' success on the EOCT in biology (Wakeman, 2013). Although the focus of the study was the predictive value of ACCESS on EOCT in biology, it is important to note that most ELLs (76%, $n = 124$) failed the EOCT in biology. In other words, three out of four students did not have enough language and content knowledge to pass the final test. The high correlation between the literacy subscale score and the EOCT in biology scores supports the need for additional research on LTELLs, the most beleaguered and emblematic ELL subgroup. If their aptitude in BICS can be converted to CALP proficiency, the students will begin to succeed in content assessments and content courses (Cummins, 1981a; Heritage et al., 2015; Johnson, 2009; Olsen, 2014; Zwiers, 2014).

In another study, researchers with the Georgia Department of Education (2008) explored the relationships between ACCESS proficiency levels and ELA EOCT scale scores. The aim of the study was to help educators correctly enroll ELL students in ninth-grade literature and composition courses and 10th-grade American literature and composition courses. The researchers sought to determine which language domain was most predictive of students' success in those courses (Georgia Department of Education, 2008). Student success was measured by scale scores on the ELA EOCT for each course. The researchers compared the scores of 962 ELLs and discovered a moderately strong and positive relationship between the scaled scores for 10th-grade American literature,

ninth-grade literature and composition EOCT, and the overall composite ACCESS scores. When exploring the predictive power of each language domain, all four domains contributed positively to the EOCT scale scores. The ACCESS reading subscale score, however, emerged as the most significant predictor of student success in the ninth-grade literature and composition and 10th-grade American literature and composition courses. The ACCESS speaking subscale scores were the least predictive of success in either course (Georgia Department of Education, 2008).

Results from the Georgia Department of Education (2008) study revealed information on the level of English language proficiency that students should demonstrate to be potentially successful in the ninth- and 10th-grade English literature and composition courses. In a range from 200 to 600, a 400 was required to pass the ELA EOCTs. The researcher of the study recommended ELLs achieve an overall composite score of 4.8 on the ACCESS to be placed in a literature and composition course, and a 4.3 to be placed in an American literature and composition course. Findings indicating that 484 (of 600) was the highest score that an ELL student achieved on the ninth-grade literature and composition EOCT and 474 (of 600) was the highest score on the American literature and composition EOCT. This finding demonstrated that ELLs who scored a 6.0, the highest level of English fluency on the ACCESS, were not scoring in the upper range of the EOCT scaled score (Georgia Department of Education, 2008). The relevance of the information gleaned from this study is that ACCESS subscale scores can be used to predict ELL students' future success in high school ELA courses. It remains unclear

whether and how their scores on the ELA EOCT translate to teacher determined course grades.

Parker et al. (2009) investigated the extent to which the four language domains of the ACCESS assessment predicted ELL performance on the reading, writing, and mathematics state content assessments. The researchers used data from the ACCESS test and the New England Common Assessment Program, which were administered to fifth (n = 1,345) and eighth (n = 921) grade ELL students in New Hampshire, Rhode Island, and Vermont. Results demonstrated the ACCESS scores were significant predictors of content assessment outcomes. Specifically, the reading and writing subscale scores proved to be significant predictors of the New England Common Assessment Program reading, writing, and mathematics scores in the fifth and eighth grades. ACCESS subscale scores on the speaking and listening sections were significant predictors of the New England Common Assessment Program scores for four outcomes: (a) fifth and eighth grade reading (speaking), (b) eighth grade writing (speaking and listening), and (c) fifth grade mathematics (listening). Findings from Parker et al.'s study were similar to those of other researchers (Flores, Batalova, & Fix, 2012; Georgia Department of Education, 2008; Wakeman, 2013) who reported that ACCESS measurements of literacy skills (encompassing reading and writing) were better predictors of content outcomes than ACCESS measurements of English oral skills (encompassing speaking and listening).

In their study related to ELL language proficiency growth and attainment, Cook and Zhao (2011) explored how many students starting at a specified level of English language proficiency attained fluent English proficiency within 5 years. The researchers used ACCESS test scores to measure students' English language proficiency. Cook and Zhao found two-thirds of the students who started at English language proficiency Level 4 attained fluent English proficiency within 5 years. Only 10% of the students who started at English language proficiency Level 1 attained a fluent English proficiency score within 5 years. The researchers concluded that students' starting levels in English language proficiency may affect the time it takes them to attain English language proficiency.

Flores et al. (2012) analyzed the educational performance trajectories of ELLs in Texas to compare their performance in school with that of their English-speaking peers. The researchers also examined whether the number of years a student was enrolled in ELL services could predict his or her academic achievement and trajectory, including high school graduation and college entry. One group sample (referred to as the on-time cohort) consisted of students (n = 133,698) who entered first grade in 1995 and advanced through to graduation with no gaps or interruptions in their schooling. This group included all students (n = 24,566) who had ever been classified as ELLs (ever-ELLs) and those who had never been (non-ELLs). Flores et al.'s findings revealed that ever-ELLs who demonstrated fluent English proficiency after 3 years were more likely to meet Texas math and reading proficiency standards than their English-only counterparts. The

researchers also found that LTELLs in the on-time cohort who remained classified as ELLs for 5 or more years lagged behind academically in every grade level. LTELLs classified as ELLs for 7 or more years fared worse. For example, 86% of 11th-grade ELL students who exited ELL programs after 3 years met math standards (Flores et al., 2012). Only 56% of LTELLs classified as ELLs for 5–6 years met math standards, whereas only 44% of LTELLs classified for 7 or more years met math standards. The researchers did not attempt to analyze or explain the LTELL results, but referred to findings from other researchers that attributed LTELLs' poor performance to "poor academic and literacy skills in English and their native language, which in turn hinder their school performance" (Flores et al., 2012, p. 13). The findings shed light on how the amount of time students spend classified as ELL students may affect their potential to meet standards, even if they remain in school until the 12th grade. Results also revealed LTELLs were less likely to enroll in 2- or 4-year institutions. The low passing scores of LTELLs raised concerns about their chances of remaining in school, graduating, and pursuing postsecondary education.

Summary

In Chapter 2, I examined several factors surrounding the topics of LTELL language proficiency, language assessments, and academic achievement. First, I discussed the characteristics of LTELLs, defined LTELLs, and examined literature pertaining to LTELLs. Researchers identified several issues related to the misidentification of LTELLs as special education students (Artiles et al., 2005; Cheatham

et al., 2014; Reeves, 2006; Rhodes et al., 2005; Sullivan, 2011), the identification and reclassification of LTELLs (Bailey & Carroll, 2015; Estrada & Wang, 2013; Linquanti & Bailey, 2014; Slama, 2012; Olsen, 2010), and U.S. educational practices and policies that inadvertently hindered LTELLs' attainment of academic English proficiency and educational success (Bailey & Carroll, 2015; Bailey & Wolf, 2012; Olsen, 2014). Next, I focused on the important role that second language acquisition plays in the development of LTELLs. I provided an overview of language proficiency, discussed the importance of academic language, and used Krashen's (1982) and Cummins's (1979) theories to explain the relevance of second language acquisition on the instruction and academic success of LTELLs. Finally, I reviewed academic grading (Cross & Frary, 1996; Marzano, 2000; Oosterhof, 2001) and the role that ELPAs (Linquanti & Bailey, 2014) have in predicting student success.

As the population of LTELLs in U.S. schools continues to grow (Heritage et al., 2015; Olsen, 2010), the need for proper identification and assessment of their academic learning and English language skills increases. Research is needed to determine if language proficiency assessments are indicative of academic progress. The purpose of this quantitative study was to examine to what extent English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment predict semester course grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students. Results from this investigation highlight variations in the efficacy of existing English courses to address the needs of LTELLs, as well as the utility of the

ACCESS for evaluating and predicting students' learning in the core academic English courses. Such findings may ultimately affect the funding, staffing, and curricula of English programs that focus on increasing the rate of academic language development, the achievement scores, and the graduation rates of LTELLs.

This chapter included a detailed analysis and synthesis of existing studies related to the research. In the following chapter, I provide a comprehensive discussion of the study's methodology. Chapter 3 includes a presentation of the research questions guiding the study, the research design and rationale, the methodology, the threats to validity, and a chapter summary. The section on methodology details the population, the procedures for sampling and recruitment, as well as participation, data collection, instrumentation, operationalization of constructs, and the data analysis plan.

Chapter 3: Research Method

The purpose of this quantitative study was to examine if, and to what extent, English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment predicted semester course grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students. I used a quantitative, nonexperimental, research design to examine the relationship between the English language proficiency scores of LTELLs on the ACCESS and their course semester grades in high school English 9, English 10, and English 11. I conducted ordinal logistic regression analyses to address each research question. This chapter contains the research questions guiding the study, the research design and rationale, the methodology, the threats to validity, and a chapter summary. In the section on methodology, I cover population, procedures for sampling and recruitment, participation, data collection, instrumentation, operationalization of constructs, and the data analysis plan.

Research Design and Rationale

I used a quantitative, nonexperimental design to investigate the predictive relationship between English language proficiency scores on the ACCESS assessment and course semester grades in high school English courses. The study was longitudinal because it followed the same cohort of students during a 3-year period. Students with complete data sets were the only students included in the analysis. A data set consisted of Spring 2012–2014 speaking, listening, reading, and writing ACCESS scores, and English 9, English 10, and English 11 first and second semester letter grades from the 2012–2013

through the 2014–2015 school years. The criterion variable for the study was academic performance, operationalized as students' fall and spring course semester letter grades in English 9, English 10, and English 11. Data analysis included all students who met the inclusion criteria and had ACCESS scores and English grades in all years. The predictor variables were the students' language proficiency scores on the speaking, listening, reading, and writing language domains of the ACCESS assessment. I conducted a total of six ordinal logistic regression analyses. These six regressions comprised two regressions for each academic year—one analysis for fall semester grades per year and one analysis for spring semester grades per year.

I selected a quantitative approach for this study because I intended to collect measurements on numerical data and use this data to test statistical relationships, as suggested by Howell (2013) and Tabachnick and Fidell (2012). Researchers may opt to use a quantitative, qualitative, or mixed methods approach to address their research questions. The choice of design is contingent on the type of research questions, procedures, and analysis the researcher intends to conduct (Field, 2013). A quantitative approach allows the testing of objective hypotheses by examining the relationships among variables (Creswell, 2009). The quantitative approach was appropriate for my study to examine the predictive relationship between students' English language proficiency ACCESS scores and their course semester grades in core academic high school English courses using statistical techniques.

This study followed a nonexperimental design using regression analysis. A correlational approach allowed for examination of the relationship between academic performance and English language proficiency (George & Mallery, 2010). Field (2013) posited that hypotheses may be tested by observation of what happens naturally, or one can manipulate some aspect of the environment and observe the effect it has on the variable(s) of interest. Nonexperimental techniques facilitate investigation of relationships when there is no control of some of the research variables and manipulation is not possible (Lappe, 2000). Correlational techniques also allow researchers to view what occurs naturally without influencing what happens nor biasing the measures of the variables with their presence (Field, 2013). This study was constrained by my inability to manipulate the groups for student classification and the lack of an intervention under investigation. Because of this, a nonexperimental approach was instrumental in overcoming the constraints imposed upon the study. Nonexperimental research does not provide information about the contiguity between different variables. Also, within nonexperimental research, it is difficult to avoid and control confounding variables. Confounding variables may make it difficult to interpret results (Field, 2013).

Methodology

Population

Servicing 325,032 students in 357 schools, the WDSD is in a burgeoning metropolis and ranks sixth in the nation in terms of the number of ELLs enrolled (WDSD, 2017). The District serviced 59,234 ELLs, 17,365 of whom were LTELLs. It

was estimated that more than 4,555 of those students also received special education services (WDSD, 2017). Similar to the rest of the nation, the WDSD struggled to meet the challenge of educating and graduating students who had not attained English language proficiency (Heritage et al., 2015; Olsen, 2010, 2014; WDSD, 2016).

Participants

In this study, I examined longitudinal trajectories of academic English proficiency for a cohort of WDSD LTELLs from 2012–2015. I investigated the degree to which their scores on the speaking, listening, reading, and writing language domains of the ACCESS assessment predicted their course semester letter grades in high school English 9, English 10, and English 11. The study participants comprised students from 53 public high schools within the WDSD and included all LTELL students who were enrolled in the ninth grade in 2012–2013 and subsequently enrolled in two additional years of school: 2013–2014 and 2014–2015. I adopted a census approach, which entailed collecting the data pertaining to all members of the available LTELL 2012–13 ninth-grade cohort. Although I did not intend to generalize findings to a larger population, a census approach is useful in producing benchmark data for future studies.

The original database file included a sample size of 718 LTELL students. To select the students to include in the cohort, the available student information (gender, race or ethnicity, receipt of special education services, and reclassification status) was extracted from the WDSD's Infinite Campus database system. Table 2 presents frequencies and percentages of gender and ethnicity for this group. Male participants

comprised more than half of the student population (*n* = 443, 62%). Most LTELL students in the population were of Hispanic origin (*n* = 675, 94%).

Table 2

Frequencies and Percentages for Gender and Ethnicity Within Selected Population (N = 718)

Variable	*n*	%
Gender		
Female	275	38
Male	443	62
Race/ethnicity		
Hispanic	675	94
Caucasian	17	2
Asian	15	2
Black	4	<1
Native American/Alaska Native	3	<1
Native Hawaiian, Other Pacific Islander	2	<1
Multiracial	2	<1

The original sample size included 718 students, however students with incomplete data sets (*n* = 218, 30%) were not included in the analysis. A data set consisted of speaking, listening, reading, and writing ACCESS scores on successive tests administered in the first semester of the 2012, 2013, and 2014 school years and fall first and spring second semester letter grades in English 9, English 10, and English 11 courses from the 2012–2013 through the 2014–2015 school years.

Furthermore, students identified as special education students (*n* = 218, 30%) were not included in the analysis. Students enrolled in one or more Honors English 9, English 10, or English 11 course semesters (*n* = 65, 9%) or those who repeated one or

more English 9, English 10, or English 11 course semesters (n = 24, 3%) were also excluded from the study. Additionally, I assessed 32 outliers and subsequently removed them from the data set. In total, 78% (n = 557) of my original sample size was excluded from the analysis. The final sample size group consisted of 161 LTELLs.

Sampling and Sampling Procedures

Because I used a predetermined student cohort, the sample was one of convenience. Convenience sampling is a common nonprobability sampling method (Frankfort-Nachmias & Nachmias, 2015). Frankfort-Nachmias and Nachmias (2015) made a distinction between probability and nonprobability sampling and described the major designs associated with each. The three major types of nonprobability samples are convenience samples, purposive samples, and quota samples. Purposive sampling entails selecting sampling units that, in the researcher's opinion, are representative of the population. Quota sampling entails the selection of a sample that is as similar as possible to the sampling population. Because my sampling strategy entailed using a population that was conveniently available and not specifically gathered to be representative of the population, purposive sampling and quota sampling were not accurate descriptions of the sampling method. The five major types of probability sampling, which include simple random sampling, systemic sampling, stratified sampling, cluster sampling , and multistage sampling were equally inapplicable because they involve compiling a desired sample by some type of random selection (Frankfort-Nachmias & Nachmias, 2015).

Although convenience sampling potentially introduced some biases into the study, I considered this possibility largely mitigated because I utilized a census sample and thus accessed and used available data for every available member of the target population who met the inclusion criteria. I accessed data available on LTELLs through the school district that employed me. I also used archival data from 2012–2015 related to ELPA scores, ELL designation, ELL status, and course semester grades in English 9, English 10, and English 11. This participant information was extracted from the WDSD's Infinite Campus database. As previously stated, the sample consisted of all LTELL language learners who met the inclusion criteria. The study participants were students from 53 public high schools within the WDSD and included the entire group of ninth-grade students who were categorized as LTELLs during the 2012–2013 school year and then persisted in the WDSD for three consecutive school years. Only students with complete data sets were included in the analysis. A complete data set consisted of Spring 2012–2014 speaking, listening, reading, and writing ACCESS scores, and English 9, English 10, and English 11 first and second semester letter grades from the 2012–2013 through the 2014–2015 school years. Students with complete data sets who repeated one or more English 9, English 10, or English 11 course semesters were excluded from the study. LTELLs from the initial ninth-grade cohort who were identified as special education, left the district, were not enrolled in English courses, or who missed at least one ACCESS assessment from 2012–2015 were also excluded from the study.

I used G*Power 3.1.9.2 to calculate the necessary sample size for the statistical analysis (Heinrich Heine Universität Düsseldorf, 2016). Each research question was addressed using ordinal logistic regression analysis. Six ordinal logistic regressions were conducted to determine if speaking, listening, reading, and writing language proficiency scores on the ACCESS assessment predict LTELLs' fall course first semester and spring second semester grades in English 9, English 10, and English 11.

Howell (2013) recommended that to determine needed sample size, power (1-β) be set to .80, alpha set to .05, and effect size designated as small. Creswell (2009) added that an effect size "identifies the strength of the conclusions about group differences or the relationships among variables in qualitative studies" (p. 167). I used the power of .80, instead of the default of .95, because it is considered the standard and most commonly used to determine sample size before conducting a data analysis (Field, 2013; Trochim, 2006). The sample size for this study was calculated using the parameters set forth by Lipsey and Wilson (2001). Using the calculations recommended by Lipsey and Wilson, I conducted a power analysis for a logistic regression to determine a sufficient sample size. Using an alpha of .05, a power of .80, a medium effect size (odds ratio = 1.72), and a two-tailed test, the minimum sample necessary to achieve statistical validity was 177 participants (Lehmann, 2006).

Procedures

After the WDSD's Department of Research and the Institutional Review Board (IRB) approved the research, I retrieved archival data spanning 2012–2015 from the

district's Infinite Campus database. The archival data set included each student's Spring 2012–2014 speaking, listening, reading, and writing ACCESS scores, and English 9, English 10, and English 11 first and second semester letter grades from the 2012–2013 through the 2014–2015 school years. The content course semester letter grades in English measured the criterion variable, student performance. This information was reported as a nominal variable (i.e., A, B, C, D, and F). The ACCESS scores on the language domains of listening, speaking, reading, and writing represented the predictor variables. The data set also included data related to ELL designation, ELL status, and school enrollment information. I used the school enrollment information to determine students' LTELL classification, which was based on the number of years students were enrolled in a U.S. school without obtaining English language proficiency.

ACCESS

The WIDA Consortium developed and launched ACCESS in 2005. ACCESS is a large-scale, high stake, standards-based and criterion-referenced English language proficiency test. As of 2015, 37 state education agencies adopted ACCESS and administered it to more than 1,372,611 ELL students annually in K–12 classrooms throughout the United States (WIDA, 2016). The WDSD first implemented ACCESS during the 2012–2013 school year (WDSD, 2016). ACCESS replaced the Language Assessment Scales Links. Students classified as ELLs in the WDSD are required to take the ACCESS assessment each year until they reach English language proficiency. The test assesses students' abilities in speaking, listening, reading, writing, and

comprehending the English used in language arts, mathematics, science, and social studies courses, in accordance with WIDA English language proficiency standards (WIDA Consortium, 2015). The technical report, *Development and Field Test of ACCESS for ELLs* (2006), demonstrated that the ACCESS domains of English fluency (listening, reading, writing, and speaking), although thematically related, are independent of each other (Kenyon, 2006). I verified the independent nature of these variables by assessing multicollinearity. I assessed the absence of multicollinearity through examination of Variance Inflation Factors (VIFs). The absence of multicollinearity assumes that the predictor variables are not too related or interdependent.

Student results in the ACCESS assessments are reported as raw scores, scale scores, and English language proficiency scores. Raw scores indicate the number of items or tasks a test taker responded to correctly out of the total number of items or tasks on the test. Scale scores make it possible to compare raw scores across grades and tiers within any of the of the four language domains on a single vertical scale. Scale scores range from 100 to 600. Scale scores represent grade levels in relation to the continuum of language development, allowing state educational agencies to compare student progress across grades, from K–12. English Language Proficiency (ELP) scores range from 1–6 and each bears a descriptive participle: 1-Entering, 2-Beginning, 3-Developing, 4-Expanding, 5-Bridging, and 6-Reaching (WIDA Consortium, 2015). Because scale scores make it possible to compare test scores across grades and tiers, within any of the four language domains, I used the scores in this study to determine whether students' socially-

acquired fluency in English enhanced or impeded their subsequent ability to internalize the cognitive skills required for progressive academic success.

Construct Reliability and Validity

The WIDA Consortium implemented measures to ensure high content reliability and validity on their ACCESS assessments. The ninth (2014) and 10th (2015) annual ACCESS technical reports provided detailed information on the construction and measurement procedures for the ninth (Series 301) and 10th (Series 302) editions of the test administered by the WDSD from 2012–2014. Series 301 was administered during the 2012–2013 academic year to 1,236,415 students in 31 states (WIDA Consortium, 2014a). Series 302 was administered during the 2013–2014 academic year to 1,372,806 students in 33 states (WIDA Consortium, 2015). The reports evidenced the reliability of test scores. The reports provided information on the accuracy and consistency of proficiency level classifications and described the scaling and equating of test forms. Also included in the reports were item-level analyses across item difficulty levels, the fit of items to the Rasch measurement model, and differential item functioning analysis for each assessment task (WIDA Consortium, 2014a, 2015).

The WDSD used the overall composite proficiency score of the ACCESS assessment to make decisions based on gains in student proficiency, to program future courses, and to terminate language support services for students who test as *English language proficient* (WDSD, 2016; WIDA Consortium, 2014a, 2015). The overall composite proficiency score is based on a student's performance in the speaking,

listening, reading, and writing test components and is weighted as follows: speaking (15%), listening (15%), reading (35%), and writing (35%; WIDA Consortium, 2014a, 2015). According to the WIDA guiding principles of language development, the weighted scheme is supported by the theory- and research-based perspective that students develop language proficiency in these four areas interdependently, at different rates, and in different ways (Gottlieb & Hamayan, 2007; Spolsky, 1989; Vygotsky, 1962). Reading and writing are weighted more heavily because they are considered to be in the academic domain, whereas listening and speaking skills theoretically start before the child is enrolled in school and are strengthened by daily use, both in and out of class (WIDA Consortium, 2015). Psychometric results indicated the reliability of the overall composite scores was very high across all grade-level clusters in both the ninth and 10th series of the ACCESS assessment. The ninth series indicated reliability scores of .973 for kindergarten; .943 for Grades 1–2; .931 for Grades 3–5; .925 for Grades 6–8; and .943 for Grades 9–12 (WIDA Consortium, 2014a). The 10th series indicated reliability scores of .973 for Kindergarten; .943 for Grades 1–2; .937 for Grades 3–5; .930 for Grades 6–8; and .945 for Grades 9–12 (WIDA Consortium, 2015). The composite scores across grades demonstrated excellent reliability for the ACCESS instrument (George & Mallery, 2010). Therefore, the ACCESS was proven to be a reliable instrument to gauge English language proficiency.

In addition to information on reliability, the reports also include information on validity. The *Standards for Educational and Psychological Testing* (hereafter referred to

as *Standards*) define validity as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9). The *Standards* also conceptualize validity as "accumulating evidence to provide a scientifically sound argument to justify the intended interpretations of test scores" (Wolf et al., 2008, p. 9). From this perspective, test validation entails compiling evidence to justify certain uses of a test rather than validating the test itself (Wolf et al., 2008).

According to the *Standards*, the validation process entails collecting evidence to support validity claim arguments. Content validity is one of several types of evidence that can be collected. Content validity can be obtained by examining how well the content of the assessment represents the construct being measured by the test. Both curricular validity and face validity studies can be used to determine content validity of a test (American Educational Research Association et al., 2009; Wolf et al., 2008). A second type of evidence is criterion-related validity. This evidence entails examining the relationships between test scores and outcomes. A criterion-related validity study can be useful in predicting concurrent or future behaviors (American Educational Research Association et al., 2009). Evidence of predictive validity, for example, shows the extent to which the ACCESS assessment scores positively predict future performance, such as the academic success of a student in a course.

A third type of evidence is construct validity. This type of evidence relates to the structure of the test and the extent to which test items and components coincide with how the construct is defined. Both convergent and discriminant validity are needed to demonstrate construct validity. Evidence of convergence, for example, refers to how well test scores and other measures of the same construct are highly correlated. Evidence of discrimination compares test scores with measures of different constructs to determine the extent to which they are not highly correlated (American Educational Research Association et al., 2009; Wolf et al., 2008). The final type of evidence is related to test takers and their response process; the researcher examines the extent that the processes students use in responding to test items and tasks match what the test is supposed to measure (American Educational Research Association et al., 2009; Wolf et al., 2008). According to Wolf et al. (2008), fairness or bias is an additional component of a test's validity that researchers should examine and analyze.

The *Standards* evidence-centered approach to validity grounds WIDA's CALS Validation Framework. The CALS is a seven-step framework that details claims related to the validity of the ACCESS assessment (WIDA Consortium, 2015). The claims are presented as a series of statements that connect some facets of the ACCESS assessment process to the intended purposes of the assessment. Evidence for each claim is then organized by the action used to support it and includes results from test data analysis, outside documentation, and other resources (WIDA Consortium, 2015). The WIDA conducts ongoing validation studies and has validated the following claims,

1. ACCESS appropriately measures different levels of social and academic English proficiency for ELLs.
2. Test takers of ACCESS are classified appropriately according to the proficiency levels defined in the WIDA English language development standards.
3. ACCESS for ELLs measures English language proficiency for all test takers in a fair and unbiased manner.
4. The same scale scores obtained by test takers of ACCESS in different years retain the same meaning.
5. ACCESS test items/tasks appropriately measure each test taker's English language proficiency.
6. All ACCESS test items and tasks are scored consistently for all test takers.
7. All test takers of ACCESS are provided comparable opportunities to demonstrate their English language proficiency (WIDA, 2015).

In their extensive research on validation studies of ELPAs, Wolf et al. (2008) found WIDA to have the most complete and thoroughly documented validation work available to the public. The WIDA validation framework outlines how validity is established by considering consequences, decisions, interpretations, assessment records, assessment performance, design, and planning. Wolf et al. noted that WIDA made clear interpretive arguments that explained how the evidence collected supported the use of their assessments.

Academic Grades

In the WDSD, achievement grades are expressed as letter grades. The letter grade is meant to represent the teacher's best judgment of how the student performed relative to the explicit learning objectives of the course. The WDSD reports student achievement for all subjects in secondary schools (Grades 6 to 12) and computes GPAs and class ranks based on the symbols and scales depicted in Table 3 (WDSD, 2006).

Table 3

WDSD Secondary School Grading Scale (Grades 6-12)

Letter Grade	Performance	%	GPA Weight
A	Excellent	90–100	4.0
B	Above Average	80–89	3.0
C	Average	70–79	2.0
D	Below Average	60–69	1.0
F	Failure	Below 60	0
P	Passing		

Note. Students are issued two quarter grades per semester. The quarter grades reflect student achievement on assignments and assessments provided by the teacher. Both quarters must be equally weighted (WDSD, 2006).

Muñoz and Guskey (2015) suggested that judgments about what students produce should be meaningful and accurate. The researchers recommended that both the purpose of grades and the format used to report them should be addressed to make grades more meaningful. That is, the instrument used to measure student learning should be valid and reliable. Per WDSD district regulations, it is expected that teachers (a) teach to objectives as outlined by state standards and district curriculum, (b) assess what it is that they

genuinely want to measure, and (c) determine grades in relationship to those objectives and assessments in a fair and consistent manner (WDSD, 2016).

Since the 2012–2013 academic year, the WDSD has implemented professional learning communities to ensure that teachers uphold district regulations and provide grades that are valid and reliable indicators of student achievement. WDSD English teachers are required to participate in professional learning communities in which they are paid up to 6 hours a month to meet and develop common homework assignments, common assessments, and common grading practices for their content area. Professional learning leaders from each high school content area attend district-wide trainings. The content leaders return to their schools and share information from district-wide trainings with their respective content area learning community team members.

Race, Brown, and Smith (2005) argued that for assessments to be valid, they should assess what educators want to measure. Race (2009) also stated reliability is synonymous with consistency and fairness and felt that assessing the work of students fairly and reliably was the most important thing that educators could do for learners. The researcher proposed interrater marking (in which teachers mark student's work and come to a unified decision on a rating scale) as a method of achieving grading system reliability. Through district regulations and professional learning communities, the WDSD implements measures to ensure that grades are valid and reliable indicators of student achievement (WDSD, 2016).

Data Analysis

I analyzed the quantitative data in this study using ordinal logistic regression analysis. All data were uploaded into SPSS Version 24 and screened for missing values, outliers, and accuracy. I examined the data set for any nonrandom patterns in missing data. There were no missing data in the data set. I screened outliers through the examination of standardized scores. Standardized scores higher than 3.29 units from the sample mean were considered evidence of an outlier (Tabachnick & Fidell, 2012). Data points with a *z* score higher than 3.29 were removed. I summarized descriptive statistics, including frequencies and percentages for categorical variables and means and standard deviations for continuous variables. These values were used to screen for inaccuracies in the data set. Values that fell outside of the expected range of values were removed.

The following research questions and hypotheses guided my study. Each question was addressed using ordinal logistic regression analysis.

RQ1. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2012–2013 assessment predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$1a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment do not predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a1a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment predict course first semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ2. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2012–2013 assessment predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$2a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–-2013 assessment do not predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a2a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment predict course second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ3. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2013–2014 assessment predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$3a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment do not predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a3a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment predict course first semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ4. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2013–2014 assessment predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$4a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–-2014 assessment do not predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a4a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment predict course second semester grades in English 10 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ5. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2014–2015 assessment predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$5a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment do not predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a5a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment predict course first semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

RQ6. To what extent, if any, can English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2014–2015 assessment predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students?

$H_0$6a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment do not predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

H_a6a. Speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment predict course second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students.

I conducted six ordinal logistic regressions to determine if speaking, listening, reading, and writing language proficiency scores on the ACCESS assessment predicted LTELLs' fall course first semester and spring course second semester grades in English 9, English 10, and English 11. I used SPSS Version 24 for the data analysis.

I assessed the assumptions of ordinal logistic regression prior to conducting the analysis. For the ordinal logistic regression to function properly, the criterion variable must be ordinal (Tabachnick & Fidell, 2012). Academic performance was assessed using course semester letter grade data, which was an ordinal, ranked measure. I conducted one

ordinal logistic regression analysis for each fall and spring course semester letter grade for English 9, English 10, and English 11. The analysis included a total of six ordinal logistic regressions. The assumptions of linearity and absence of multicollinearity were assessed for the logistic regression analysis. The assumption of linearity in ordinal logistic regression requires a linear relationship between any continuous predictors and the logit of the outcome variable (Tabachnick & Fidell, 2012). The absence of multicollinearity was assessed through examination of Variance Inflation Factors. The absence of multicollinearity assumes that the predictors are not too related or interdependent. Variance Inflation Factor values higher than 10 indicate that the assumption of multicollinearity was violated (Stevens, 2009). The proportional odds assumption, which assumes the relationship between all groupings is the same, can be assessed using the parallel lines test. I did not conduct the test of parallel lines to assess proportional odds because of the failure of the nonproportional odds to converge.

I used the Nagelkerke R^2 to assess the contribution of each independent variable to the variation in the dependent variable. The χ^2 coefficient was used to assess statistical significance for the overall model consisting of the selected independent variables. The Wald coefficient expressed the influence of the individual predictor variables. Estimated *B* was reported to represent the likelihood of an increase in the outcome category given a change in a corresponding individual independent variable.

Threats to Validity

Validity refers to the degree to which correct inference can be made regarding the findings (Campbell & Stanley, 1963). Validity addresses both external and internal concerns. External validity refers to the extent to which the results of a study can be generalized confidently to a group larger than the group that participated in the study (Campbell & Stanley, 1963). Internal validity refers to the extent to which the results obtained from the research study are a function of the variables that were systematically manipulated, observed, or measured in the study. Potential threats to internal validity include history, maturation, testing, instrumentation, statistical regression, mortality, and selection (Campbell & Stanley, 1963). Fraenkel and Wallen (2012) suggested four general ways to maximize internal validity:

1. Standardizing the conditions under which the research study is carried out will help minimize threats to internal validity from history and instrumentation.
2. Obtaining as much information as possible about the participants in the research study will help minimize threats to internal validity from mortality and selection.
3. Obtaining as much information as possible about the procedural details of the research study will help minimize threats to internal validity from history and instrumentation.
4. Choosing an appropriate research design can help control most other threats to internal validly.

In correlational studies, in which no interventions occur, threats to internal validity, such as implementation, history, maturation, attitude or subjects, and regression threats, are not applicable (Fraenkel & Wallen, 2012). There are some threats, however, related to subject characteristics, mortality, location, instrument decay, testing, history, data collector characteristics, and data collector biases that do apply.

A possible internal validity threat to the study was location. A location threat is possible whenever instruments are not administered to participants under the same conditions (Fraenkel & Wallen, 2012). The administration of ACCESS in the WDSD occurred within a 6-week testing window that spanned the months of February to March each year. The ACCESS assessment was administered at 53 different school locations. The administrators of each school had the liberty to select how, when, and where they administered the ACCESS. In some high schools, for example, participants may have been provided a comfortable, well-lit, and well-ventilated room in which to test. In other high schools, participants may have been placed in noisy, uncomfortable, poorly lit rooms. The events that transpired within each school location could have altered the outcomes or results of the study and hence produced a misleading correlating coefficient. Although this potential threat existed, WDSD schools were directed to maintain the integrity of a consistent environment for testing, which minimized this threat.

A second possible threat to internal validity included instrumentation, specifically data collector characteristics. Data collector characteristics can create a threat if different persons administer the instruments (Fraenkel & Wallen, 2012). Gender, age, or ethnicity,

for example, can affect participants' responses. The WDSD administered ACCESS to more than 68,000 students each year. School personnel at the different school locations administered the listening, reading, and writing sections. The speaking section of ACCESS, which was required to be administered individually to each student, was administered by one of the 54 district testers who were trained to administer and score the speaking section. The 54 district testers traveled to each school location to administer the speaking section of the ACCESS assessment to each student individually. The WDSD had standardized instruments, administration of data collection procedures, and required trainings to help control the instrumentation threat. Although the possibility exists that the idiosyncratic behavior of some of the 54 district data testers may have affected the outcome of the study, ongoing training and screening at the WDSD mitigated this threat. Teacher grading policies, teacher preferences, and consistency in grading posed additional threats to internal validity. These threats were mitigated through the process WDSD schools utilized in establishing the instructional framework for the district and the grading schema described in the instrumentation.

A third potential threat to internal validity was testing. Testing relates to the possible effects of a first test (or instrument) on the subsequent performance by study participants on a second test (or instrument; Fraenkel & Wallen, 2012). The study did not include a pre- and posttest. It did, however, include the administration of a similar test in the 3 years examined (2012-2015). Per the WIDA Consortium's Technical Report No. 10, one-third to one-half of all ACCESS test items were refreshed each year (WIDA,

2015). Although a 1-year lapse in between ACCESS testing occurred, there exists the possibility that study participants may have recalled items from the previous test that could have affected their performance.

External validity refers to the extent to which the results of a study can be generalized confidently to a group larger than the group that participated in the study, and can be divided into ecological and population validity (Campbell & Stanley, 1963). Ecological validity is present to the degree that a result generalizes across settings. There was a threat to the ability to generalize my findings across settings, given that the WDSD and the students it served may not be identical to other districts and their student population. Population validity refers to how representative the study sample is of the population and how widely the research findings apply. Because I used all available members of the population who fulfilled the inclusion criteria within the study, threats to population validity were mitigated.

Transiency also posed a unique threat to this study regarding LTELL status, especially in a district with an overall transiency rate of 40%. Because I solely included students with complete ACCESS and English course semester letter grades for English 9, English 10, and English 11, the sample of LTELL students included in the study may have differed from the population of LTELL students in WDSD. However, the difference was minor provided that the transiency rate among LTELLs in the WDSD was just below 13% (WDSD, 2016).

Ethical Considerations

I did not have contact with participants during this study. The ethical considerations of the study were related to receipt of appropriate approvals to conduct the study and to access the data, and to the careful storage and treatment of the archival data. Additionally, I had to consider how the identity of students was safeguarded and how their confidentiality was maintained during the process of the study.

After Walden University IRB approval (approval number: 10-26-17-0070939), I requested the data file from the WDSD Department of Research. The Department of Research required that an application be completed and approved before I could obtain the required data. The district's application explicitly requested administrator sponsorship information, the name of the affiliated university, its Institutional Review Board (IRB) application and approval letter, and an Assurance Training Certificate. After I was approved to conduct the research, I requested a de-identified data file for the study. The district de-identified the data before releasing it to me, hence, there was no need to recode the data to protect student identities. No individual- or school-level identifiers were necessary for the stated purpose of this research. After I received the electronic copy of the data file, I stored it on my personal, password-protected computer. I only provided aggregate-level information in this dissertation and all public reports. The raw data were available to those directly associated with the project. The data will be stored for a period of 5 years, after which time I will scrub the data file from my computer drive.

Summary

This chapter detailed the methodology for the quantitative, nonexperimental study. For this study, I evaluated the predictive relationship between LTELLs' grades in high school English classes and the language proficiency scores that they received in the speaking, listening, reading, and writing language domains of the ACCESS assessment. I analyzed archival data for a cohort of ninth-grade students classified as LTELL learners from 2012–2015 using ordinal logistic regression analysis. The research design and rationale, the methodology, and the threats to validity were described in this chapter. I presented the population, sampling, data collection, instrumentation, and operationalization of the study's constructs, as well as my data analysis plan. Chapter 4 will include the results from the analysis of the data.

Chapter 4: Results

The purpose of my study was to examine the predictive relationship between English language proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment and first and second course semester grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students. I analyzed archival data using ordinal logistic regression analyses. Six research questions were posed to measure the predictive relationship between speaking, listening, reading, and writing on the ACCESS assessment for 2012–2015, and semester course grades in English 9, English 10, and English 11.

This chapter contains a description of the data collection for my study. I present the preliminary data management steps and the findings of the ordinal logistic regression analyses in the results section. I close the chapter with a summary of the salient aspects of the data analysis.

Data Collection

In this study, I examined longitudinal trajectories of academic English proficiency for a cohort of 718 ninth-grade WDSD LTELLs from 2012–2015. I investigated the degree to which their scores on the speaking, listening, reading, and writing language domains of the ACCESS assessment predicted their course semester letter grades in high school English 9, English 10, and English 11. The study participants comprised students from 53 public high schools within the WDSD and included all LTELL students who were enrolled in the ninth grade in 2012–2013 and subsequently enrolled in two

additional years of school: 2013–2014 and 2014–2015. I adopted a census approach, which entailed collecting the data pertaining to all members of the available LTELL 2012–2013 ninth-grade cohort. Although I did not intend to generalize findings to a larger population, a census approach was useful in producing benchmark data for future studies.

The original database file obtained from the WDSD included a sample size of 718 LTELL students. At this point, I sorted the data file and removed students with incomplete data sets (n = 218, 30%) and students identified as special education students (n = 218, 30%). Additionally, students with complete data sets who repeated one or more English 9, English 10, or English 11 course semesters were also removed (n = 65, 9%), as well as students with complete data sets who were enrolled in one or more honors English courses (n = 24; 3%). Table 4 presents frequencies and percentages of gender and ethnicity for the 193 participants who comprised the sample. Male participants comprised more than half of the student sample (n = 112, 58%). All LTELL students in the sample size group were of Hispanic origin (n = 193, 100%).

Table 4

Frequencies and Percentages for Gender and Ethnicity Within Selected Population (N = 193)

Variable	*n*	%
Female	81	42
Male	112	58
Hispanic	193	100

I entered the archival data file into SPSS Version 24 for data management and analysis. The file contained data for 193 LTELL students from the selected district. I screened the data file for nonrandom patterns in missing data. There were no entries in the data set with missing data; therefore, no cases were removed from the data set. I then screened the data set for outliers within the continuous ACCESS subscale scores by calculating standardized scores for the variables. Data points with a standardized score higher than ±3.29 were considered outliers (Tabachnick & Fidell, 2012). I subsequently removed 32 outliers from the data set. For the 2012–2013 ACCESS scores, three outliers existed in the listening scores, three outliers in the speaking scores, and one outlier in the writing scores. For the 2013–2014 ACCESS scores, 10 outliers existed in listening scores and three outliers in writing scores. For the 2014–2015 ACCESS scores, 10 outliers existed in reading scores and two outliers in speaking scores.

I calculated descriptive statistics for the ACCESS data and the semester grades to screen for accuracy. All values were within the range of acceptable values. Therefore, I determined that no inaccurate values existed in the data set. The final data set contained data for 161 cases. I conducted a post hoc power analysis to assess the achieved power, given that the final number of participants in the sample was lower than the minimum sample size (n = 177) needed for statistical validity. For the ordinal logistic regression with an odds ratio of 1.72, an alpha of 0.05, and 161 cases, the achieved power was 0.76 (Heinrich Heine Universität Düsseldorf, 2016).

Results

I calculated the frequencies and percentages of the English course grades for the first and second semesters in Grades 9, 10, and 11. Table 5 presents descriptive statistics for the English course grades. C was the most common English course grade for English 9 during the first semester (n = 53, 33%), English 9 second semester (n = 54, 34%), English 10 first semester (n = 55, 34%), English 10 second semester (n = 55, 34%), and English 11 first semester (n = 53, 33%). Responses varied for English 11 for the second semester; however, many students received a D (n = 47, 29%).

Table 5

Descriptive Statistics for English Course Grades for First and Second Semester, Grades 9-11

	First Semester		Second Semester	
Variable	*n*	*%*	*n*	*%*
English 9				
A	5	3	7	4
B	31	19	22	14
C	53	33	54	34
D	41	25	43	27
F	31	19	35	22
English 10				
A	11	7	4	2
B	33	21	26	16
C	55	34	55	34
D	34	21	46	29
F	28	17	30	19
English 11				
A	7	4	4	2
B	32	20	25	16
C	53	33	41	25
D	38	24	47	30
F	31	19	44	27

Note. Due to rounding errors, percentages may not equal 100%.

Next, I calculated means and standard deviations for ACCESS scores in listening, reading, speaking, and writing for each academic year. Table 6 presents the means and standard deviations for 2012–2013 ACCESS scores. During that academic year, the lowest ACCESS mean score was reading ($M = 365.60$, $SD = 14.78$), and the highest ACCESS mean score was speaking ($M = 405.21$, $SD = 27.47$).

Table 6

Means and Standard Deviations for ACCESS Scores, 2012–2013 Academic Year

Variable	*M*	*SD*	*n*
Listening	383.97	24.38	161
Reading	365.60	14.78	161
Speaking	405.21	27.47	161
Writing	401.02	17.70	161

Table 7 presents the means and standard deviations for 2013–2014 ACCESS scores. During that academic year, the lowest ACCESS mean score was reading (M = 372.25, SD = 16.50). The highest ACCESS mean score was the writing score (M = 403.06, SD = 15.17).

Table 7

Means and Standard Deviations for ACCESS Scores, 2013–2014 Academic Year

Variable	*M*	*SD*	*n*
Listening	397.25	25.45	161
Reading	372.25	16.50	161
Speaking	387.74	31.83	161
Writing	403.06	15.17	161

Table 8 presents the means and standard deviations for 2014–2015 ACCESS scores. During the 2014–2015 academic year, the ACCESS reading mean score was

lowest (M = 374.61, SD = 18.12). The ACCESS writing score was highest (M = 402.55, SD = 14.79).

Table 8

Means and Standard Deviations for ACCESS Scores, 2014–2015 Academic Year

Variable	*M*	*SD*	*n*
Listening	402.13	31.19	161
Reading	374.61	18.12	161
Speaking	401.32	29.80	161
Writing	402.55	14.79	161

Finally, I reported the means, standard deviations, and associated proficiency levels for students' grades by academic year and semester (Tables 9–11). I provided this information to increase understanding of students' achievement on the ACCESS, as indicated by the mean score of letter grade in the course. This information was vital because the ACCESS subscales have different score ranges for proficiency levels. The six levels of language proficiency identified by ACCESS scores are, 1-Entering, 2-Beginning, 3-Developing, 4-Expanding, 5-Bridging, and 6-Reaching.

Table 9 presents mean, standard deviation, and proficiency level ACCESS subscale scores for English Grade 9 first and second semester. Student ACCESS proficiency levels ranged from Levels 1 or 2 to Levels 5 or 6 across student grade levels and course semesters. Students achieved the highest proficiency level on ACCESS speaking. Students who received a C grade achieved a mean proficiency level of 5 or 6

for both semesters, students who received a D grade achieved a mean proficiency level of 5 or 6 in speaking for the second semester, and students who received an F grade achieved a mean proficiency level of 5 or 6 in speaking for both semesters. Students achieved the lowest mean proficiency level on the reading subscale, with students who received an F grade achieving a proficiency level of 1 or 2 on the reading subscale during first semester. For the remaining grade levels and semesters, students who received an A, B, C, D or F achieved a mean proficiency level of 2 or 3 for the reading subscale. Students who received a D grade also received a listening subscale score of 2 or 3 during the second semester, a decline from the 3 or 4 mean proficiency level that they achieved during the first semester.

During Grade 10, students' achieved proficiency levels ranged from Levels 2 or 3 to Levels 4 or 5. The highest mean proficiency level was achieved during first semester on the speaking subscale by students who received a B. However, the mean proficiency level for students who received a B decreased from a Level 4 or 5 to a Level 3 or 4 during second semester on the speaking subscale. There was an increase from first to second semester among students who received an F from a Level 2 or 3 to a Level 3 or 4 on the writing subscale. Similar to Grade 9, Grade 10 students achieved the lowest mean proficiency level on the reading subscale. Overall, the mean proficiency levels achieved by students on the speaking component of the ACCESS decreased from Grade 9 to Grade 10. Table 10 presents means, standard deviations, and proficiency level ACCESS subscales during first and second semester for English 10.

Table 9

ACCESS Proficiency Level by English 9 Grade

Grade	ACCESS Subscale	First Semester			Second Semester		
		M	*SD*	Proficiency Level	*M*	*SD*	Proficiency Level
A	12/13 Listening	383.20	12.44	3/4	389.00	15.42	3/4
	12/13 Reading	371.80	10.03	2/3	371.14	9.58	2/3
	12/13 Speaking	383.80	41.03	3/4	396.43	39.84	4/5
	12/13 Writing	403.40	19.24	3/4	402.71	16.34	3/4
B	12/13 Listening	394.13	22.48	3/4	393.18	18.49	3/4
	12/13 Reading	366.42	11.98	2/3	369.23	14.70	2/3
	12/13 Speaking	404.55	29.47	4/5	404.55	27.28	4/5
	12/13 Writing	408.19	13.18	3/4	407.86	11.90	3/4
C	12/13 Listening	382.45	23.80	3/4	388.02	25.70	3/4
	12/13 Reading	366.15	14.43	2/3	365.43	13.90	2/3
	12/13 Speaking	407.49	26.41	5/6	402.69	31.26	5/6
	12/13 Writing	401.94	14.37	3/4	402.44	18.79	3/4
D	12/13 Listening	383.15	28.69	3/4	377.40	27.52	2/3
	12/13 Reading	365.27	16.20	2/3	364.42	14.53	2/3
	12/13 Speaking	403.29	28.19	4/5	407.28	23.95	5/6
	12/13 Writing	395.90	20.16	3/4	397.77	15.79	3/4
F	12/13 Listening	377.61	20.28	2/3	379.00	20.33	2/3
	12/13 Reading	363.29	16.85	1/2	363.94	17.23	2/3
	12/13 Speaking	407.97	23.75	5/6	408.74	23.13	5/6
	12/13 Writing	398.68	21.24	3/4	398.20	20.65	3/4

Table 10

ACCESS Proficiency Level by English 10 Grade

Grade	ACCESS Subscale	First Semester			Second Semester		
		M	*SD*	Proficiency Level	*M*	*SD*	Proficiency Level
A	13/14 Listening	403.18	23.21	3/4	399.25	14.75	3/4
	13/14 Reading	370.55	13.74	2/3	371.50	19.57	2/3
	13/14 Speaking	389.36	35.31	3/4	388.00	27.47	3/4
	13/14 Writing	412.27	9.79	3/4	412.75	14.59	3/4
B	13/14 Listening	399.88	28.92	3/4	400.81	19.37	3/4
	13/14 Reading	376.03	10.30	2/3	375.69	11.47	2/3
	13/14 Speaking	394.24	32.15	4/5	387.00	34.05	3/4
	13/14 Writing	407.82	12.36	3/4	409.27	10.21	3/4
C	13/14 Listening	396.51	24.36	3/4	399.73	28.85	3/4
	13/14 Reading	372.73	19.81	2/3	372.29	18.19	2/3
	13/14 Speaking	383.36	30.35	3/4	390.20	30.32	3/4
	13/14 Writing	404.31	15.25	3/4	405.95	14.82	3/4
D	13/14 Listening	401.24	22.55	3/4	395.09	24.01	3/4
	13/14 Reading	370.38	15.24	2/3	369.98	17.11	2/3
	13/14 Speaking	387.44	35.84	3/4	384.26	35.33	3/4
	13/14 Writing	399.74	18.86	3/4	398.65	16.89	3/4
F	13/14 Listening	388.46	26.72	3/4	392.70	27.00	3/4
	13/14 Reading	369.82	17.86	2/3	372.80	16.08	2/3
	13/14 Speaking	388.39	28.52	3/4	389.17	28.84	3/4
	13/14 Writing	395.39	10.37	2/3	397.83	13.58	3/4

During Grade 11, students' mean proficiency levels ranged from Levels 1 or 2 to Levels 5 or 6 on the ACCESS subscales. On the reading subscale, students who received

an A or a D in the class declined from a Level 2 or 3 to a Level 1 or 2 from first to second semester. However, C students increased from a Level 1 or 2 to a Level 2 or 3 on the reading subscale from first to second semester. Students who received an A in the class also declined from first to second semester on the speaking subscale, moving from a Level 5 or 6 to a Level 4 or 5. Students who received an A, B, C, or F had a mean proficiency of 4 or 5 or a level of 5 or 6 on the speaking subscale for first and second semester. Table 11 presents the means, standard deviations, and proficiency levels on the ACCESS subscales for English 11.

Table 11

ACCESS Proficiency Level by English 11 Grade

Grade	ACCESS Subscale	First Semester			Second Semester		
		M	*SD*	Proficiency Level	*M*	*SD*	Proficiency Level
A	14/15 Listening	416.57	19.40	4/5	421.50	6.35	4/5
	14/15 Reading	383.29	13.76	2/3	370.75	8.02	1/2
	14/15 Speaking	419.71	21.92	5/6	410.50	35.00	4/5
	14/15 Writing	412.57	8.16	3/4	413.50	5.80	3/4
B	14/15 Listening	409.88	23.37	3/4	412.88	19.83	3/4
	14/15 Reading	380.38	12.96	2/3	381.76	13.78	2/3
	14/15 Speaking	400.28	28.10	4/5	405.88	30.95	4/5
	14/15 Writing	408.00	10.12	3/4	404.72	12.31	3/4
C	14/15 Listening	406.74	24.71	3/4	407.34	22.57	3/4
	14/15 Reading	372.98	18.86	1/2	379.88	13.43	2/3
	14/15 Speaking	399.96	32.48	4/5	400.83	30.20	4/5
	14/15 Writing	403.62	13.76	3/4	405.95	14.90	3/4
D	14/15 Listening	397.39	20.14	3/4	398.51	23.03	3/4
	14/15 Reading	374.53	20.09	2/3	371.51	18.13	1/2
	14/15 Speaking	396.00	29.08	3/4	396.91	31.03	3/4
	14/15 Writing	396.50	17.90	2/3	401.19	16.27	2/3
F	14/15 Listening	388.81	51.26	3/4	393.27	46.19	3/4
	14/15 Reading	369.58	18.47	1/2	369.30	22.23	1/2
	14/15 Speaking	407.06	28.26	4/5	403.05	27.60	4/5
	14/15 Writing	400.23	14.68	2/3	398.59	13.96	2/3

To address the research questions, I conducted six ordinal logistic regressions, two per academic year corresponding to first semester and second semester course grades. I assessed the predictive relationship between ACCESS speaking, listening, reading, and writing scores and semester grades for English 9, English 10, and English 11. Within the data set, I coded a semester score of an A as a 1, a B as a 2, a C as a 3, a D

as a 4, and an F as a 5. I present the results of each analysis, organized by null hypothesis. Negative coefficients indicated higher semester grades while positive coefficients indicated lower semester grades according to the coding schema.

$H_0$1a. 2012–2013 ACCESS Scores Do Not Predict First Semester English 9 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factors values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10; therefore, the assumption was met (see Table 12). The test of parallel lines was not conducted to assess proportional odds because of the failure of the nonproportional odds to converge. Because of the small sample sizes within groups and the failure of the nonproportional odds to converge, I advise the reader to exercise caution in interpreting the results.

The results of the analysis, $\chi^2(4) = 10.77$, $p = .029$, indicated that the model containing 2012–2013 Listening, Reading, Speaking, and Writing and predicting English 9 first semester grades was statistically significant. Because of this finding, I rejected the null hypothesis. I calculated the McFadden's R-squared to assess model fit, with values higher than .2 indicating excellent fit (Louviere, Hensher, & Swait, 2000). Values below the 0.2 to 0.4 range for the McFadden's R-squared are considered evidence of poorer model fit, with lower estimates evidencing extremely poor fit (McFadden, 1973). The McFadden R-squared value for this model was 0.02, indicating poor model fit. The poor model fit may be attributed to other contributing variables or latent variables that have

not been included in the model. However, because the model was deemed statistically significant despite the poor model fit, I assessed the contribution of the individual predictors to the variation in the criterion variable. Additionally, the presence of statistical significance indicates that a statistically significant trend does exist between the predictors and the criterion variable, while the low McFadden R-squared indicates that there is significant variability in the criterion variable to which the predictors contribute little (Field, 2013; Hensher & Swait, 2000).

The individual predictors were assessed for their contribution to the variation in English 9 first semester grades. The regression coefficient for 2012–2013 Listening was not significant, $B = -0.01$, $\chi^2 = 2.71$, $p = .100$, indicating the listening scores did not predict English 9 first semester grades. The regression coefficient for 2012–2013 Reading was not significant, $B = -0.01$, $\chi^2 = 0.30$, $p = .584$, indicating the reading scores did not predict English 9 first semester grades. The regression coefficient for 2012–2013 Speaking was not significant, $B = 0.00$, $\chi^2 = 0.74$, $p = .391$, indicating the speaking scores did not predict English 9 first semester grades. However, the regression coefficient for 2012–2013 Writing was significant, $B = -0.02$, $\chi^2 = 4.31$, $p = .038$. This finding indicated 2012–2013 Writing scores did predict the criterion variable. The odds ratio for the writing score was 0.98, indicating a small decrease in likelihood that students would do poorly in the course with an increase in their writing score. The results of the analysis are presented in Table 12.

Table 12

Ordinal Logistic Regression Results for 2012–2013 ACCESS Scores predicting English 9 First Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.01	0.01	2.71	.100	0.99	1.05
Reading	-0.01	0.01	0.30	.584	0.99	1.07
Speaking	0.00	0.01	0.74	.391	1.00	1.01
Writing	-0.02	0.01	4.31	.038	0.98	1.07

**Note*. $\chi^2(4) = 10.77$, $p = .029$.

$H_0$2a. 2012–2013 ACCESS Scores Do Not Predict Second Semester English 9 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factor values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10 therefore the assumption was met (see Table 13). I conducted the test of parallel lines to assess proportional odds. The test assesses the null hypothesis that the slope coefficients are identical across categories of the dependent variable. I failed to reject the null hypothesis; therefore, the assumption was met, $\chi^2(12) = 6.04$, $p = .914$.

The results of the analysis, $\chi^2(4) = 13.25$, $p = .010$, indicated that the model containing 2012–2013 Listening, Reading, Speaking, and Writing and predicting English 9 second semester grades was statistically significant. Because of this finding. I rejected

the null hypothesis. I calculated a McFadden's R-squared of 0.03 for the model, indicating poor model fit.

The individual predictors were assessed for their contribution to the variation in English 9 second semester grades. The regression coefficient for 2012–2013 Listening was significant, $B = -0.01$, $\chi^2 = 4.87$, $p = .027$, indicating that the listening scores did contribute to the variation in English 9 second semester grades. = .038. This finding indicated that 2012–2013 Listening scores did predict the criterion variable. The odds ratio for the listening score was 0.99, indicating a small decrease in likelihood that students would do poorly in the course with an increase in their listening score. The regression coefficient for 2012–2013 Reading was not significant, $B = -0.01$, $\chi^2 = 0.87$, $p = .351$, which indicated that the reading scores did not predict English 9 second semester grades. The regression coefficient for 2012–2013 Speaking was not significant, $B = 0.01$, $\chi^2 = 2.24$, $p = .134$, which suggested that the speaking did not predict the criterion variable. The regression coefficient for 2012–2013 Writing was not significant, $B = -0.01$, $\chi^2 = 2.47$, $p = .116$, which indicated that the predictor did not predict English 9 second semester grades. The results of the analysis are presented in Table 13.

Table 13

Ordinal Logistic Regression Results for 2012–2013 ACCESS Scores predicting English 9 Second Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.01	0.01	4.87	.027	0.99	1.05
Reading	-0.01	0.01	0.87	.351	0.99	1.07
Speaking	0.01	0.01	2.24	.134	1.01	1.01
Writing	-0.01	0.01	2.47	.116	0.99	1.07

**Note*. $\chi^2(4) = 13.25$, p = .010.

$H_0$3a. 2013–2014 ACCESS Scores Do Not Predict First Semester English 10 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factor values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10; therefore, the assumption was met (see Table 14). The test of parallel lines assesses the null hypothesis that the slope coefficients are identical across categories of the dependent variable. The test of parallel lines was not conducted to assess proportional odds because of the failure of the nonproportional odds to converge. Because of the small sample sizes within groups and the failure of the nonproportional odds to converge, I advise the reader to exercise caution in interpreting the results.

The results of the analysis, $\chi^2(4) = 16.46$, $p = .002$, indicated that the model containing 2013–2014 Listening, Reading, Speaking, and Writing predicting English 10 first semester grades was statistically significant. Because of this finding, I rejected the

null hypothesis. The McFadden R-squared value calculated for this model was 0.03, indicating poor model fit.

Because the model was statistically significant, I assessed the contribution of the individual predictors to the variation in English 10 first semester grades. The regression coefficient for 2013–2014 Listening was not significant, $B = -0.01$, $\chi^2 = 1.12$, $p = .290$, indicating that listening scores did not predict English 10 first semester grades. The regression coefficient for 2013–2014 Reading was not significant, $B = 0.00$, $\chi^2 = 0.05$, $p = .828$, indicating that reading scores did not predict English 10 first semester grades. The regression coefficient for 2013–2014 Speaking was not significant, $B = -0.00$, $\chi^2 = 0.01$, $p = .917$, indicating speaking scores did not predict English 10 first semester grades. The regression coefficient for 2013–2014 Writing was significant, $B = -0.04$, $\chi^2 = 12.10$, $p < .001$. This finding suggested that the writing score predicted the likelihood of receiving a poor grade in the course. The odds ratio was 0.96, indicating that the likelihood of receiving a poor grade in the course decreased with an increase in students' writing score. The results of the analysis are presented in Table 14.

Table 14

Ordinal Logistic Regression Results for 2013–2014 ACCESS Scores predicting English 10 First Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.01	0.01	1.12	.290	0.99	1.14
Reading	0.00	0.01	0.05	.828	1.00	1.21
Speaking	-0.00	0.00	0.01	.917	1.00	1.06
Writing	-0.04	0.01	12.10	< .001	0.96	1.08

**Note*. $\chi^2(4) = 16.46$, $p = .002$.

$H_0$4a. 2013–2014 ACCESS Scores Do Not Predict Second Semester English 10 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factor values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10 therefore the assumption was met (see Table 15). I conducted the test of parallel lines to assess proportional odds. The test assesses the null hypothesis that the slope coefficients are identical across categories of the dependent variable. The null hypothesis was not rejected, $\chi^2(12) = 4.73$, $p = .967$; therefore, the assumption was met.

The results of the analysis, $\chi^2(4) = 15.24$, $p = .004$, indicated that the model consisting of 2013–2014 Listening, Reading, Speaking, and Writing and predicting English 10 second semester grades was statistically significant. Because of this finding, I

rejected the null hypothesis. The McFadden R-squared for this model was 0.03, indicating poor fit.

I assessed the individual predictors because the model was statistically significant. The regression coefficient for 2013–2014 Listening was not significant, $B = -0.01$, $\chi^2 = 1.13$, $p = .288$, which indicated that listening scores did not predict English 10 second semester grades. The regression coefficient for 2013–2014 Reading was not significant, $B = 0.00$, $\chi^2 = 0.20$, $p = .658$, which indicated that reading scores did not predict English 10 second semester grades. The regression coefficient for 2013–2014 Speaking was not significant, $B = 0.00$, $\chi^2 = 0.05$, $p = .824$, which indicated that speaking scores did not predict English 10 second semester grades. The regression coefficient for 2013–2014 Writing was significant, $B = -0.04$, $\chi^2 = 11.60$, $p < .001$. This finding suggested that 2013–2014 writing scores did predict the criterion variable. The odds ratio was 0.97, indicating that an increase in the writing score decreased the likelihood that students would receive a poor grade in the class. The results of the analysis are presented in Table 15.

Table 15

Ordinal Logistic Regression Results for 2013–2014 ACCESS Scores predicting English 10 Second Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.01	0.01	1.13	.288	0.99	1.14
Reading	0.00	0.01	0.20	.658	1.00	1.21
Speaking	0.00	0.00	0.05	.824	1.00	1.06
Writing	-0.04	0.01	11.60	< .001	0.97	1.08

**Note*. $\chi^2(4) = 15.24$, p = .004.

$H_0$5a. 2014–2015 ACCESS Scores Do Not Predict First Semester English 11 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factor values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10; therefore, the assumption was met (see Table 16). I conducted the test of parallel lines to assess proportional odds. The test assesses the null hypothesis that the slope coefficients are identical across categories of the dependent variable. The null hypothesis was not rejected, $\chi^2(12) = 17.82$, $p = .121$; therefore, the assumption was met. However, because of the small sample sizes within groups and the failure of the nonproportional odds to converge, I advise the reader to exercise caution in interpreting the results.

The results of the analysis, $\chi^2(4) = 23.78$, $p < .001$, indicated that the model consisting of 2014–2015 Listening, Reading, Speaking, and Writing and predicting English 11 first semester grades was statistically significant. Because of this finding, I

rejected the null hypothesis. The McFadden R-squared value for this model was 0.05, indicating poor model fit.

I assessed the contribution of the individual predictors to the variation in English 11 first semester grades. The regression coefficient for 2014–2015 Listening was significant, $B = -0.02$, $\chi^2 = 7.11$, $p = .008$. The odds ratio was 0.98, indicating that as the listening score increased the likelihood of receiving a poor grade in English 11 during first semester decreased. The regression coefficient for 2014–2015 Reading was not significant, $B = -0.01$, $\chi^2 = 1.24$, $p = .265$, which indicated that reading scores did not predict English 11 first semester grades. The regression coefficient for 2014–2015 Speaking was not significant, $B = 0.00$, $\chi^2 = 0.11$, $p = .738$, which suggested that speaking scores did not predict English 11 first semester grades. The regression coefficient for 2014–2015 Writing was significant, $B = -0.03$, $\chi^2 = 7.92$, $p = .005$. The odds ratio for writing was 0.97, indicating that the likelihood of receiving a poor grade in English 11 during first semester decreased as the writing score increased. The results of the analysis are presented in Table 16.

Table 16

Ordinal Logistic Regression Results for 2014–2015 ACCESS Scores predicting English 11 First Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.02	0.01	7.11	.008	0.98	1.10
Reading	-0.01	0.01	1.24	.265	0.99	1.15
Speaking	0.00	0.00	0.11	.738	1.00	1.05
Writing	-0.03	0.01	7.92	.005	0.97	1.02

**Note.* $\chi^2(4) = 23.78$, $p < .001$.

$H_0$6a. 2014–2015 ACCESS Scores Do Not Predict Second Semester English 11 Grades

Prior to conducting the analysis, I assessed the assumptions of multicollinearity and proportional odds. I calculated and screened VIFs to assess multicollinearity. Variance Inflation Factor values higher than 10 were considered evidence of multicollinearity. None of the VIF values exceeded 10; therefore, the assumption was met (see Table 17). The test of parallel lines assesses the null hypothesis that the slope coefficients are identical across categories of the dependent variable. The test of parallel lines was not conducted to assess proportional odds because of the failure of the nonproportional odds to converge. Because of the small sample sizes within groups and the failure of the nonproportional odds to converge, I advise the reader to exercise caution in interpreting the results.

The results of the analysis, $\chi^2(4) = 21.11$, $p < .001$, indicated that the model consisting of 2014–2015 Listening, Reading, Speaking, and Writing and predicting

English 11 second semester grades was statistically significant. Because of this finding, I rejected the null hypothesis. The McFadden R-squared value for this model was 0.05, indicating poor model fit.

Because the model was statistically significant I assessed the contribution of the individual predictors to the variation in English 11 second semester grades. The regression coefficient for 2014–2015 Listening was significant, $B = -0.01$, $\chi^2 = 4.83$, $p = .028$. The odds ratio was 0.99, indicating a minimal decrease in the likelihood a student would receive a poor grade in English 11 during the second semester. The regression coefficient for 2014–2015 Reading was significant, $B = -0.02$, $\chi^2 = 4.23$, $p = .040$. The odds ratio for reading was 0.98. This finding indicated a small decrease in the likelihood that a student would receive a poor grade in English 11 during second semester with an increase in the reading score. The regression coefficient for 2014–2015 Speaking was not significant, $B = 0.00$, $\chi^2 = 0.01$, $p = .922$, which indicated that speaking scores did not predict English 11 second semester grades. The regression coefficient for 2014–2015 Writing was significant, $B = -0.02$, $\chi^2 = 3.92$, $p = .048$. The odds ratio for writing was 0.98. This finding indicated a small decrease in the likelihood that students would receive a poor grade in English 11 during second semester with an increase in their writing score. It is important to note that the *p*-value was close to .05, indicating marginal significance. The results of the analysis are presented in Table 17.

Table 17

Ordinal Logistic Regression Results for 2014-2015 ACCESS Scores predicting English 11 Second Semester Grades

Predictor	*B*	*SE*	χ^2	*p*	*OR*	*VIF*
Listening	-0.01	0.01	4.83	.028	0.99	1.10
Reading	-0.02	0.01	4.23	.040	0.98	1.15
Speaking	0.00	0.00	0.01	.922	1.00	1.05
Writing	-0.02	0.01	3.92	.048	0.98	1.02

**Note.* $\chi^2(4) = 21.11$, $p < .001$.

Summary

The purpose of this study was to examine the predictive relationship between English language proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment and course semester grades in English 9, English 10, and English 11 for a cohort of 718 ninth-grade LTELL students. I analyzed archival data using ordinal logistic regression analyses. The ordinal logistic regression analyses were conducted to assess the relationship between ACCESS scores and LTELL students' first and second semester English course grades in Grades 9, 10, and 11. The results of the analysis indicated that during the 2012–2013 academic year, ACCESS writing scores predicted English 9 first semester grades, while listening scores predicted English 9 second semester grades. During the 2013–2014 academic year, ACCESS writing scores predicted English 10 first semester and English 10 second semester grades. During the 2014–2015 academic year, ACCESS listening and writing scores predicted English 11 first semester grades, while listening, reading, and writing predicted English 11 second

semester grades. For each predictor variable, a one-unit increase in the predictor decreased the likelihood of receiving a lower grade in the course.

I presented the data collection and preliminary data management steps in this chapter. I covered the results of the data analysis, organized by research question. In Chapter 5, I will provide a summary and interpretation of the findings and describe the limitations of the study. I will also present recommendations for future research and practice and describe the implications for positive social change.

Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this quantitative study was to examine if, and to what extent, English proficiency subscale scores in speaking, listening, reading, and writing on the ACCESS assessment could predict semester course grades in English 9, English 10, and English 11 for a 2012–2015 cohort of ninth-grade LTELL students. I conducted ordinal logistic regression to evaluate these predictive relationships. By examining the predictive value of the ACCESS assessment scores on LTELLs' English course semester grades, this study provided insight into whether oral proficiency scores are used appropriately in the WDSD and elsewhere, and whether other subscales (listening, reading, and writing) are better predictors of LTELL academic success.

ACCESS subscale scores comprised the predictor variables and English course semester grades comprised the criterion variables. Letter Grades A–F were used to assess LTELLs' academic performance in English 9, English 10, and English 11 courses. Because ACCESS scores and academic achievement in English classes are quantifiable variables, the nature of this study was quantitative, and I employed a nonexperimental design. Because I did not control, manipulate, or alter the predictor variable or subjects, a nonexperimental design was appropriate.

The results of my analysis indicated that during the 2012–2013 academic year, LTELLs' ACCESS writing scores predicted their English 9 first semester grades, while listening scores predicted their English 9 second semester grades. During the 2013–2014 academic year, LTELLs' ACCESS writing scores predicted their English 10 first and

second semester grades. During the 2014–2015 academic year, LTELLs' ACCESS listening and writing scores predicted their English 11 first semester grades, while listening, reading, and writing predicted their English 11 second semester grades. For each predictor variable, a 1-unit increase in the predictor decreased the likelihood of receiving a lower grade in the English course. The results also demonstrated that LTELLs' ACCESS speaking scores did not predict their English course grades at any grade level.

Interpretations of the Findings

Sample

In this study, I examined longitudinal trajectories of academic English proficiency for a cohort of ninth-grade LTELLs. The participants comprised students from 53 public high schools within the WDSD and included all LTELL students who were enrolled in the ninth grade in 2012–2013 and subsequently enrolled in 2 additional years of school: 2013–2014 and 2014–2015. The original database file obtained from the WDSD included a sample size of 718 LTELL students.

Data from WDSD indicated that only 2% of LTELLs were reclassified as English proficient in 2017. This coincided with the percentage of 2012–2013 cohort participants that were excluded from this study based on reclassification. During the 3-year span of this study, only 19 (2.6%) of the 718 study participants obtained English language proficiency. Researchers (Cook et al., 2011; Cummins, 1981b; Hakuta et al., 2000; Thomas & Collier, 2002; Zhao & Maina, 2015) have found that the average ELL takes 5–

7 years to acquire grade-level academic English language skills. Researchers contended that middle- and high-school students often take longer to progress from intermediate levels to full English language proficiency than elementary school students do (Cook et al., 2008; Kieffer, 2008, 2010, 2011). My findings supported this by demonstrating that 97.4% of LTELL cohort students had not acquired full levels of English proficiency even after being enrolled in the WDSD for 8 or more years.

In this study, 30% (n = 218) of LTELL cohort participants from the original data set were classified as special education students. This number coincided with Brunner's (2012) data, indicating that 33% of high school LTELLs were placed in special education programs.

Researchers Cheatham et al. (2014) and Sullivan (2011) found that ELL students are at risk of being overrepresented in special education programs because of inappropriate referrals (Cheatham et al., 2014; Sullivan, 2011). The research presented by Artiles et al. (2005), Sullivan (2011), and Reeves (2006) demonstrated that instructors who do not understand ELLs' second language difficulties, often mislabel them as learning disabled. In addition, teachers who hold negative views of ELLs often refer them for special education services (Cheatham et al., 2014; Harklau, 2000; Sharkey & Layzer, 2000; Walker et al., 2004). The existing research and the high number of LTELLs classified as special education students in the WDSD warrants further investigation into whether these students were appropriately identified and classified as special education students.

The original database file obtained from the WDSD included a sample size of 718 LTELL students. The data file was sorted and students with incomplete data sets (n = 218, 30%) and students identified as special education students (n = 218, 30%) were removed. Additionally, students with complete data sets who repeated one or more course semesters of English 9, English 10, or English were also removed (n = 65, 9%), as were students with complete data sets who were enrolled in one or more honors English courses (n = 24, 3%). The final analytic sample consisted of 193 participants. Male participants comprised more than half of the student sample (n = 112, 58%). All LTELL students in the sample size group were of Hispanic origin (n = 193, 100%).

I entered the archival data file containing 193 LTELL students into SPSS Version 24 for data management and analysis. I screened the data file for nonrandom patterns in missing data. There were no entries in the data set with missing data; therefore, no cases were removed from the data set. I then screened the data set for outliers within the continuous ACCESS subscale scores by calculating standardized scores for the variables. Data points with a standardized score higher than ± 3.29 were considered outliers (Tabachnick & Fidell, 2012). I assessed 32 outliers that I subsequently removed from the data set. The final data set contained data for 161 cases. Given the original sample size of 718 LTELL cohort students, this diminution was unanticipated.

Research Questions

This study was conducted to examine the predictive relationship between English language proficiency subscale scores on the ACCESS assessment and semester grades in

English 9, English 10, and English 11. The purpose of the study was to obtain insight into whether oral proficiency scores are used appropriately in the WDSD and elsewhere, and whether other subscales (listening, reading, and writing) are better predictors of LTELLs' academic success. The WDSD data demonstrated a 66% dropout rate among LTELLs; therefore, understanding whether LTELLs' English language proficiency is a factor influencing their academic grades is important.

Research Questions 1 and 2 examined the relationship between LTELLs' English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2012–2013 assessment and their course first and second semester grades in English 9. The null hypothesis that speaking, listening, reading, and/or writing subscale scores on the ACCESS 2012–2013 assessment do not predict course first and second semester grades in English 9 for the 2012–2015 cohort of ninth-grade LTELL students was rejected. When individual predictors were assessed for their contribution to the variation in English 9 first and second semester grades, only the writing scores were found to predict English 9 first semester grades and only the listening scores were found to predict English 9 second semester grades. In studies related to ELLs, researchers (Parker et al., 2009; Wakeman, 2013) reported that ACCESS subscale scores in writing were better predictors of content outcomes than ACCESS subscale scores in reading, speaking, and listening. Therefore, increasing curricular emphasis for LTELLs on writing during the first semester and on listening during the second semester can increase their likelihood of earning higher grades in English 9.

Research Questions 3 and 4 addressed whether English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2013–2014 assessment could predict course first and second semester grades in English 10 for the same cohort of LTELL students. The null hypothesis that speaking, listening, reading, and/or writing subscale scores on the ACCESS 2013–2014 assessment do not predict course first and second semester grades in English 10 for the 2012–2015 cohort of LTELL students was rejected. When individual predictors were assessed for their contribution to the variation in English 10 first and second semester grades, only the writing scores were found to predict English 10 first and second semester grades. The findings confirmed that writing continued to predict students' grades in the second semester of English 10 and reaffirmed the need to continue developing students' writing proficiency well into 10th grade. LTELLs' oral proficiency, which was thought to predict students' academic achievement in English courses, had no influence on their English 10 first or second semester grades. These findings have the potential to diminish the importance placed on LTELLs' oral language development to guide instructional decisions made on behalf of these students.

Research Question 5 and 6 addressed whether English proficiency subscale scores in speaking, listening, reading, and/or writing on the ACCESS 2014–2015 assessment could predict first and second semester grades in English 11 for the same cohort of LTELL students. The null hypothesis that speaking, listening, reading, and/or writing subscale scores on the ACCESS 2014–2015 assessment do not predict course first and

second semester grades in English 11 for the 2012–2015 cohort of ninth-grade LTELL students was rejected. When I assessed individual predictors for their contribution to the variation in English 11 first and second semester grades, writing and listening scores were found to predict LTELLs' English 11 first semester grades and writing, reading, and listening were found to predict their second semester grades.

The results of this study demonstrated that as students moved from English 9 to English 11, the number of language domains affecting their course grades increased. Although the reading subscale scores did not predict these students' English 9 and English 10 grades, the reading subscale scores began to predict their grades as they moved on to the second semester of their English 11 course. The conclusion can be drawn that writing subscale scores predict LTELLs' English 9, English 10 and English 11 grades, and that it would be worth increasing the emphasis in that area for LTELLs across all grades. In contrast, students' oral subscale scores, which drive the instructional decisions made by WDSD instructional leaders on behalf of LTELLs, do not predict LTELLs' academic grades in English 9, English 10, or English 11. Thus, WDSD and other districts should consider altering their sources of information regarding such instructional decisions.

ACCESS Reading and Writing Subscale Scores

In studies related to ELLs, other researchers (Parker et al., 2009; Wakeman, 2013) also reported that ACCESS subscale scores in writing were better predictors of content outcomes than ACCESS subscale scores in reading, speaking, and listening. When

determining the predictive value of ELLs' ACCESS test scores and their EOCT in biology, Wakeman (2013) found that the writing subscale score had the highest predictive value of the four language domains. Additionally, Parker et al. (2009), who investigated the extent to which the four language domains of the ACCESS assessment predicted ELLs' performance on state content assessments, found reading and writing subscale scores to be significant predictors of the New England Common Assessment Program reading, writing, and mathematics scores in fifth and eighth grades. Findings from the current study were also similar to those of other researchers (Flores et al., 2012; Georgia Department of Education, 2008) who reported that ACCESS measurements of literacy skills (encompassing reading and writing) were better predictors of content outcomes than ACCESS measurements of English oral skills (encompassing speaking and listening).

ACCESS Speaking Subscale Scores

The descriptive statistics demonstrated that LTELLs' mean ACCESS speaking scores were equivalent to a language proficiency level of 5 or 6 during the 2012–2013 ninth-grade school year and a level of 4 or 5 during the 2014–2015 11th-grade school year. This revealed that in Grades 9 and 11, the cohort participants possessed well-developed conversational English skills. This validated existing research that proves LTELLs possess social skills that enable them to function at high levels in social interactions (Menken et al., 2012; Olsen, 2014; Slama, 2012).

WDSD educational leaders assume that LTELLs' oral fluency on ACCESS is an accurate predictor of their ability to understand course content and perform well in classroom speaking, listening, reading, and writing tasks (Gandy, 2013). The results of the current study demonstrated that the speaking subscale scores on ACCESS did not predict LTELLs' English 9, English 10, or English 11 grades. Thus, being able to speak English proficiently may not mean a student can use academic language well enough to succeed in English 9, English 10, or English 11 courses.

Other researchers, however, have found predictive value in the speaking component of the ACCESS test. Wakeman (2013), for example, found that the speaking component contributed positively to the EOCT scores in biology. Results from the Georgia Department of Education (2008) also revealed the speaking subscale scores were an important predictor of students' success in ninth-grade literature and composition and 10th-grade American literature and composition courses. Additionally, Parker et al. (2009) found the ACCESS speaking subscale scores to be significant predictors of the New England Common Assessment Program scores for fifth and eighth grade reading and eighth grade writing. It should be noted that in those studies, the study participants were ELL students and not limited to LTELLs.

English Language Proficiency and Grades

This study provided insight regarding the predictive relationship between LTELLs' English language proficiency and the letter grades they received from their teachers in English 9, English 10, and English 11. The descriptive statistics for the ninth-

grade 2012–2013 academic year demonstrated that in the first semester, 22% and in second semester, 18% of LTELLs received an A or B in English 9. Students who received a grade of A or B were at the same mean equivalent English language proficiency level in reading (Level 2 or 3) and writing (Levels 3 or 4) as were students who received a grade of C or D. In fact, students who received a grade of C or F had a higher mean equivalent English language proficiency level in speaking than students who received a grade of A or B.

The descriptive statistics for the 10th grade 2013–2014 school year demonstrated that in the first semester, 29% and in second semester, 18% of LTELLs received an A or B in English 10. Students who received a grade of A were at same mean equivalent English language proficiency level in listening (Levels 3 or 4), reading (Levels 2 or 3), and speaking (Levels 3 or 4) as were students who received a grade of C, D, or F. Students who received a grade of C or D were at the same mean equivalent English language proficiency level in writing (Levels 3 or 4) as students who received a grade of A or B.

The descriptive statistics for the 11th grade 2014–2015 school year demonstrated that in first semester, 24% and in second semester, 18% of LTELLs received an A or B in English 11. Students who received a grade of B were at the same mean equivalent English language proficiency level in listening (Level 3 or 4) and speaking (Level 4 or 5) as were students who received a grade of C or F. Students who received a grade of D or F

dropped to a lower mean equivalent English language proficiency level in writing (Level 2 or 3) and reading (Level 1 or 2) than in previous years.

A positive finding that emerged from this study was that LTELLs are able to obtain a grade of A or B in English 9, English 10, or English 11 despite not obtaining English language proficiency in all four language domains. Cummins (1979) argued that comprehensible input is what leads students to make sense of the academic language and acquire an understanding of the content. The researcher also contended that students who are effectively instructed in a first language (L1) can successfully transfer the language skills they develop to a second language (L2) as long as they are motivated to learn the second language. Krashen (1982) suggested that teachers who understand how to create noncompetitive environments that are encouraging and rich in visual and auditory stimuli can lower students' affective filters and provide LTELLs with diverse opportunities to experience academic success. Additionally, teachers who have some theoretical and methodological knowledge about second language learners will make use of appropriate scaffolding strategies that allow students the opportunity to complete intellectually demanding tasks they encounter in content courses (Heritage et al., 2015). Hence, LTELLs who received an A or a B in English classes may have had teachers that embedded academic language in comprehensible contexts, thus making cognitively demanding content easier for LTELLs to understand.

ACCESS

Students classified as ELLs in the WDSD are required to take the ACCESS assessment each year until they reach English language proficiency. The test assesses students' abilities in speaking, listening, reading, writing, and comprehending the English used in language arts, mathematics, science, and social studies courses, in accordance with WIDA English language proficiency standards (WIDA Consortium, 2015). An important element of the ACCESS assessment is its proven alignment with academic content standards that guide instructional classroom practices delineated by each state (Cook, 2007; WIDA, 2013).

The WDSD uses ACCESS oral proficiency scores to determine LTELLs' course placement (Abedi, 2007; U.S. Department of Education, 2013). WDSD educational leaders currently use LTELLs' oral fluency scores on ACCESS as a predictor of their abilities to understand course content and perform well in classroom speaking, listening, reading, and writing tasks (Gandy, 2013). Thus, district leaders place LTELLs in mainstream classes where they are taught the regular core curriculum without compensatory language support, accommodations, or strategies to help them master course content (Olsen, 2010). The results of this study demonstrated that the speaking subscale scores on ACCESS did not predict LTELLs' English 9, English 10, or English 11 grades. Further, the study results indicated that the likelihood of the ACCESS language domains predicting LTELLs' English 9, English 10, and English 11 course

grades was minimal, supporting the need to explore other factors that may predict English grades.

Limitations of the Study

This study had several limitations that must be addressed. First, because this study was delimited by a single school district, the sample was a unique, nonprobability convenience sample, rather than a random sample. Therefore, the results may not be generalizable beyond the specific population from which the sample was drawn. This limitation applies to all studies that are not randomized control trials. However, external validity in this study was enhanced because ACCESS has been adopted by 39 states that are part of the WIDA Consortium. I chose ACCESS as the focus of this investigation because it is the most widely used English language assessment in U.S. schools, as well as the assessment used in the WDSD (WIDA, 2016). However, ACCESS is not the only ELPA used in the United States (WIDA, 2016). Thus, the findings from this study may not be useful to students and schools that do not use ACCESS.

This study was further limited by the use of ninth-grade LTELLs from 53 public high schools within the WDSD. The students in this study remained enrolled in ninth grade in 2012–2013 and subsequently enrolled 2 additional years of school: 2013–2014 and 2014–2015. However, LTELLs in the WDSD are enrolled in many different grades and vary in age. The results of this study may not accurately reflect these students.

A third limitation to this study was that student characteristics, such as LTELLs' special education status, honor course status, or repeat course status, was not assessed.

Subject characteristics threaten the design when the possibility exists that individual characteristics other than those already correlated can explain any relationships found (Fraenkel & Wallen, 2012). Therefore, these groups of LTELLs were excluded from the design to eliminate this subject characteristic threat. Thus, the study results cannot be generalized to these subgroups of LTELLs.

A fourth and major limitation of this study was the variability found within each WDSD high school as it pertains to teachers, instructional practices, and English course content that may influence students' letter grades. This limitation was further exacerbated because each student's data set consisted of two course semester grades from three different English courses with distinct curriculums. To reduce the variability of their grades, the WDSD implemented measures to ensure that grades were valid and reliable indicators of student achievement by issuing district-wide regulations and sponsoring professional learning communities (WDSD, 2016).

A fifth limitation was the use of the proportional odds model. Despite its favorable properties in application of ordinal regressions, the model has the disadvantage of often failing to converge (Tutz & Scholz, 2003). Prior to conducting data analysis, I assessed the assumption of proportional odds. The nonproportional odds failed to converge, hence the test of parallel lines was not conducted to assess proportional odds. This could have been caused by separations in the data (separations can occur because of issues with multiple X values predicting the same Y value). This decreased the validity of the results and limited my ability to draw trusted inference from the results.

The model fit was an additional limitation. In calculating the McFadden's R-squared to assess each model fit, each regression model showed a poor model fit in the analysis. Additionally, the statistically significant predictors contributed moderately to predicting the likelihood of achieving a higher or lower grade in the English courses. This may be attributed to excluding other covariates that may predict grades, such as students' attendance rates, socioeconomic status, mother tongue, and length of time in the United States. It was important to still assess the regression model because the statistically significant results indicate a significant trend between the predictor and criterion variables, while the poor model fit indicates much variation still exists in the criterion variable that cannot be attributed to the predictors.

A final limitation of the current study was the sample size. The final sample size (n = 161) was slightly below the minimum sample size necessary (n = 177) for statistical validity. The small sample size may have contributed to the weak association found between ACCESS language proficiency scores and English course grades. However, the moderate predictive value of writing, listening, and reading ACCESS subscale scores on LTELLs' grades is sufficient to suggest the value of further research in this area with more variables in a more robust model so the sample size threshold is met or surpassed.

Recommendations

The study findings contribute to the literature on LTELLs' academic success by confirming a predictive relationship between ACCESS subscale scores in reading, writing, and listening and their English course semester grades. Thus, further

investigation into these language domains is warranted. The study results demonstrated that LTELLs' ACCESS subscale speaking scores, which currently drive the instructional decisions made by the WDSD on behalf of these students, do not predict English course grades. Thus, district initiatives promoting oral discourse as a reliable sign of LTELLs' academic achievement may be more profitably focused on the language domains of listening, writing, and reading, which were more predictive of LTELLs' English course semester grades.

Recommendations for future research include WDSD educational leaders understanding what programs and instructional practices yield higher grades. The findings may then be disseminated to school leaders who can share these effective instructional practices with teachers throughout the district. Thus, it is recommended that further studies be conducted on other LTELL subgroups (e.g., LTELL special education students, LTELLs with missing ACCESS data, LTELLs in honors English courses, or LTELLs who repeated English courses) to provide insight into the ACCESS language domains that predict the English 9, English 10, and English 11 course semester grades for these subgroups of LTELL students.

Implications

This study contributes to the understanding of the predictive value of ACCESS subscale scores on English course semester grades for LTELLs in high school. Social change for LTELLs may occur based on the findings of this study, which identified the three ACCESS subscale scores of writing, listening, and reading as predictive of

students' academic achievement in English courses. Instruction that fosters the development of these language domains may be helpful in increasing LTELLs' English 9, English 10, and English 11 grades.

Understanding the predictive value of ACCESS on LTELLs' academic performance is especially important in the WDSD, where, as of 2017–2018, the sole graduation requirement is for a student to receive passing grades in all required classes. The WDSD has relied mainly on LTELLs' oral proficiency to determine their course placement. However, as the results demonstrated, being able to speak English proficiently may not mean a student can use academic language well enough to succeed in courses. In fact, fewer than 30% of LTELL study participants placed in mainstream classes without appropriate English support were successful in obtaining a grade of A or B in English 9, English 10, and English 11.

Positive Social Change

The findings confirm a need to implement changes in the manner in which the WDSD, and possibly other school districts, use LTELLs' ACCESS speaking subscale scores to determine course placement. Although LTELLs' oral fluency may be used to assist students in making meaning of what they read, write, or listen to, it should not be the sole language domain that guides all instructional decisions made on behalf of LTELLs. This study has demonstrated that LTELLs' ACCESS subscale scores in listening, reading, and writing are better predictors of their academic success in high school English courses than their scores on the ACCESS speaking subscale. Focusing on

the language domains that best predict content achievement is a matter of equity for LTELLs, because the results inform educational decisions that significantly affect their lives. This research can benefit society by helping the WDSD, and other districts, assist LTELLs in obtaining grades that allow them to graduate from high school and transition to higher education.

Conclusion

The results of this study demonstrated that LTELLs' well-developed oral skills, which currently drive the instructional decisions made by the WDSD on their behalf, do not predict the grades they receive in academic English classes. Therefore, district initiatives that focus on promoting oral discourse in hopes of increasing students' academic achievement may need to be redirected towards developing LTELLs' overall literacy skills. LTELLs in the process of becoming language proficient have the potential to perform well in high school academic English classes, as indicated by the results of this study. The ACCESS writing and reading subscale scores were found to be contributing factors to ameliorating LTELLs' English 9, English 10, and English 11 semester grades. Thus, redirecting instructional focus in high school English courses to their academic literacy may gradually decrease LTELLs' high dropout rate.

References

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

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice, 27*(3), 17–31. doi:10.1111/j.1745-3992.2008.00125.x

Aina, J. K., Ogundele, A. G., & Olanipekun, S. S. (2013). Students' proficiency in English language relationship with academic performance in science and technical education. *American Journal of Educational Research, 1*(9), 355–358. doi:10.12691/education-1-9-2

Airasian, P. W. (2000). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill.

Akhavan, N. (2006). *Help! My kids don't all speak English: How to set up a language workshop in your linguistically diverse classroom.* Portsmouth, NH: Heinemann.

Alford, B. J., & Niño, M. C. (2011). *Leading academic achievement for English language learners: A guide for principals*. Thousand Oaks, CA: Corwin.

AlHaddad, S. K., Mohamed, M., & AlHabshi, S. M. (2004). An exploratory study on English language proficiency and academic performance in the context of globalization of accounting education. *Journal of Financial Reporting and Accounting 2*(1), 55–71. doi:10.1108/19852510480000661

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 78*(5), 218–223. Retrieved from http://www.tandfonline.com/toc/vtch20/current

Allen, P., Swain, M., Harley, B., & Cummins, J. (1990). Aspects of classroom treatment: Toward a more comprehensive view of second language education. In B. Harley et al. (Eds.), *The development of second language proficiency* (pp. 57–81). Cambridge, United Kingdom: Cambridge University Press.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition*. Cambridge, United Kingdom: Cambridge University Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, *71*(3), 283–300. doi.10.1177/001440290507100305

Ascenzi-Moreno, L., Kleyn, T., & Menken, K. (2013). *A CUNY-NYSIEB framework for the education of "long-term English learners" 6–12 grades*. New York, NY: City of New York. Retrieved from http://www.academia.edu/8490424/A_CUNY-

NYSIEB_FRAMEWORK_ FOR _THE_EDUCATION_OF_LONG-TERM_ENGLISH_LEARNERS_6-12_GRADES

ATLAS. (2015). DLLS Assessment. *New America.* Retrieved February 26, 2016 from http://atlas.newamerica.org/dll-assessments

August, D., & Shanahan, T. (2006). *Developing reading and writing in second language learners: Report of the national literacy panel on language-minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York, NY: Grune & Stratton.

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

Bailey, A. B., & Carroll, P. E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education, 39*(1), 253–294. doi:10.3102/0091732X14556074

Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy, grades K–6: Building reading and academic language skills across the curriculum.* Thousand Oaks, CA: Corwin/Sage Press.

Bailey, A. L., & Kelly, K. R. (2013). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy, 27*(5), 770–804. doi:10.1177/0895904811432137

Bailey, A. L., & Wolf, M. K. (2012, January). The challenge of assessing language proficiency aligned to the common core state standards and some possible solutions. Paper presented at Understanding Language Conference, Stanford, CA. Retrieved from http://ell.stanford.edu/sites/default/files/pdf/academicpapers/08Bailey%20Wolf%20Challenges%20of%20Assessment%20Language%0Proficiency%20FINAL_0.pdf

Bloom, B. S., Engelhar, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain.* New York, NY: David McKay.

Brookhart, S. M. (1993). Teachers grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123–142. doi:10.1111/j.1745-1993.tb01070.x

Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). New York, NY: Addison Wesley Longman.

Brunner, J. (2012). *English language learners summary, 2010–2011* (Publication No. 10.59). Austin, TX: Austin Independent School District.

Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305–328. doi:10.3102/00028312042002305

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton Mifflin.

Carnock, J. T. (2016). After AMAOs: Defining what progress for English learners means under ESSA. *New America.* Retrieved from https://www.newamerica.org /education-policy/edcentral/ells-essa/

Cheatham, G. A., Jimenez-Silva, M., Wodrich, D. L., & Kasai, M. (2014). Disclosure of information about English proficiency. *Journal of Teacher Education, 65*(1) 53–62. doi:10.1177/0022487113503687

Chi, Y., Garcia, R. B., Surber, C., & Trautman, L. (2011). Alignment study between the Common Core State Standards in English language arts and mathematics and the WIDA English Language Proficiency Standards, 2007 edition, prekindergarten through grade 12. Norman, OK: The University of Oklahoma Education Training, Evaluation, Assessment, and Measurement. Retrieved from https://www.wida.us /get.aspx?id=371

Chomsky, N. (1959). Review of verbal behavior by B. F. Skinner. *Language. 35*(1), 26–58. Retrieved from https://chomsky.info/1967____/

Common Core State Standards Initiative. (n.d.). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects.* Retrieved from http://www.corestandards.org/wp-content/uploads /ELA_Standards1.pdf

Cook, H. G. (2007). *Alignment study report: The WIDA consortium's English language proficiency standards for English language learners in kindergarten through grade 12.* Retrieved from https://www.wida.us/research/agenda/alignment/

Cook, H. G. (2014). *Examining relationships between alternate ACCESS and state alternate assessments: Exploring notions of English language proficiency*. Madison, WI: Wisconsin Center for Education Research.

Cook, H. G., Boals, T., & Lundberg, T. (2011). Academic achievement for English learners: What can we reasonably expect? *Phi Delta Kappan, 93*(3), 66–69. doi:10.1177/003172171109300316

Cook, H. G., Boals, T., Wilmes, C., & Santos, M. (2008). *Issues in the development of annual measurable achievement objectives for WIDA consortium states* (WCER Working Paper No. 2008-2). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research. Retrieved from http://www.wcer .wisc.edu/publications/workingPapers/Working_Paper_No_2008_02.pdf

Cook, H. G., & Zhao, Y. (2011, April). *How English language proficiency assessments manifest growth: An examination of language proficiency growth in a WIDA state*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Cornwell, C. M., Mustard, D. B., & Van Parys, J. (2008, Jun 25). How does the new SAT predict academic achievement in college? (University of Georgia Working Paper). Athens, GA: University of Georgia. Retrieved from http://people.terry.uga.edu /mustard/New%20SAT. pdf

Crane, E. W., Barrat, V. X., & Huang, M. (2011). *The relationship between English proficiency and content knowledge for English language learner students in*

grades 10 and 11 in Utah. (Issues & Answers Report, REL 2011–No. 110). Washington, DC: U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approach* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Cross, L. H., & Frary, R. B. (1996, April). Hodgepodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY. Retrieved from http://files.eric.ed.gov/fulltext/ED398262.pdf

Cummins, J. (1979). *Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. Working papers on bilingualism, no. 19*. (Bilingual Education Project Working Paper No. 19). Toronto, Canada: Ontario Institute for Studies in Education. Retrieved from https://www.researchgate.net/publication/234573070_CognitiveAcademic_Language_Proficiency_Linguistic_Interdependence_the_Optimum_Age_Question_and_Some_Other_Matters_Working_Papers_on_Bilingualism_No_19

Cummins, J. (1981a). Age on arrival and immigrant second language leaning in Canada: A reassessment. *Applied Linguistics, 11*(2), 132–149. doi:10.1093/applin/2.2.132

Cummins, J. (1981b). The role of primary language development in promoting educational success for language minority students. In Rachel B. S. (2012). A

longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology, 104*(2), 265-285. doi:10.1037/a0025861

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. London, England: Kogan Page.

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

Cummins, J., & Hornberger, N. H. (2008). *Bilingual education*. New York, NY: Springer.

Díaz-Rico, L., & Weed, K. (2002). *The crosscultural, language, and academic development handbook: A complete K–12 reference*. Boston, MA: Allyn & Bacon.

DiCerbo, P. A., Anstrom, K. A., Baker, L. L., & Rivera, C. (2014). A review of the literature of teaching academic English to English language learners. *Review of Educational Research, 84*(3), 446–482. doi:10.3102/0034654314532695

DuFour, R., DuFour, R., Eaker, R., Many, T., & Mattos, M. (2016). *Learning by doing: A handbook for professional learning communities at work* (3rd ed.). Bloomington, IN: Solution Tree Press.

Dutro, S., & Moran, C. (2003). Rethinking English language instruction: An architectural approach. In J. Zwiers (Ed.). *Building academic language: Meeting common core standards across disciplines* (2nd ed., pp. 22). San Francisco, CA: Jossey-Bass.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, England: Oxford University Press.

Estrada, P., & Wang, H. (2013, May). Reclassifying and not reclassifying English language learners to fluent proficient, year 1 findings: Factors impeding and facilitating reclassification and access to the core. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* Millbrae, CA: California Academic Press.

Fakeye, D. (2014). English language proficiency as a predictor of academic achievement among EFL students in Nigeria. *Journal of Education and Practice 5*(9), 38–41. Retrieved from http://www.iiste.org/Journals/index.php/JEP/article /viewFile/11863/12212

Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal, 3*(4), 70–85. Retrieved from https://ehlt.flinders.edu.au /education/iej/articles/v3n4/feast/paper.pdf

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, England: Sage Publications.

Flores, S. M., Batalova, J., & Fix, M. (2012). *The educational trajectories of English language learners in Texas*. Washington, DC: Migration Policy Institute.

Forte, E., & Faulkner-Bond, M. (2010). *The administrator's guide to federal programs for English learners.* Washington, DC: Thompson.

Forte, E., Perie, M., & Paek, P. (2012). *Exploring the relationship between English language proficiency assessments and English language arts assessments*. Washington, DC: Center for Assessment, U.S. Department of Education.

Fox, J., & Fairbairn, S. (2011). ACCESS for ELLs. *Language Testing, 28*(3), 425–431. doi:10.1177/0265532211404195

Fraenkel, J. R., & Wallen, N. E. (2012). *How to design and evaluate research in education* (8th ed.). New York, NY: McGraw-Hill.

Frankfort-Nachmias, C., & Nachmias, D. (2015). *Research methods in the social sciences* (8th ed.). New York, NY: Worth.

Freeman, R. (2002). *Building on community bilingualism: Promoting multiculturalism through schooling.* Philadelphia, PA: Caslon Publishing.

Freeman, Y. S., & Freeman, D. E. (2009). *Academic language for English language learners and struggling readers*. Portsmouth, NH: Heinemann.

Gallagher, J. D. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Gándara, P., & Hopkins, M. (2010). *Forbidden language*. New York, NY: Teachers College Press.

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives, 11*, 36. doi:10.14507/epaa.v11n36.2003

Gandy, S. E. (2013). Informal reading inventories and ELL students. *Reading & Writing Quarterly, 29*(3), 271–287. doi:10.1080/10573569.2013.789782

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference, 18.0 update* (11th ed.). Boston, MA: Allyn and Bacon.

Georgia Department of Education. (2008). *What Georgia educators need to know about Georgia's testing program*. Atlanta, GA: Author.

Gewertz, C. (2007). Remediation for exit-exam failure proves daunting. *Education Week, 26*(18), 10. Retrieved from http://www.edweek.org/ew/articles/2007/01/10 /18exit.h26.html

Ghenghesh, P. (2015). The relationship between English language proficiency and academic performance of university students- should academic institutions really be concerned? *International Journal of Applied Linguistics and English Literature, 4(*2), 91–97. doi:10.7575/aiac.ijalel.v.4n.2p.91

Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, CA: Corwin Press.

Gottlieb, M. (2013). *Large-scale assessment of English language learners: Addressing educational accountability in K–12 settings.* Alexandria, VA: Teachers of English to Speakers of Other Languages.

Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity*. Thousand Oaks, CA: Sage Publications.

Gottlieb, M., Cranley, M. E., & Cammilleri, A. (2007). *Understanding the WIDA English language proficiency standards*. Madison, WI: Board of Regents of the University of Wisconsin System.

Gottlieb, M., & Ernst-Slavit, G. (2013). *Academic language in diverse classrooms: Mathematics Series*. Thousand Oaks, CA: Corwin Press.

Gottlieb, M., & Hamayan, E. (2007). Assessing oral and written language proficiency: A guide for psychologists and teachers. In G. B. Esquivel, E. C. Lopez, & S. G. Nahari (Eds.), *Handbook of multicultural school psychology: An interdisciplinary perspective* (pp. 245–263). New York, NY: Erlbaum.

Gottlieb, M., Katz, A., & Ernst-Slavit, G. (2009). *Paper to practice: Using the English language proficiency standards in PreK-12 classrooms.* Alexandria, VA: Teachers of English to Speakers of Other Languages.

Grant, R., Cook, H. G., & Phakiti, A. (2011). *Relationship between language proficiency and mathematics achievement*. Madison, WI: Board of Regents of the University of Wisconsin System.

Graves, M. (2006). *The vocabulary book: Learning and instruction*: New York, NY: Teachers College Press.

Gredler, M. E. (1999). *Classroom assessment and learning*. New York, NY: Longman.

Hakuta, K. (2014). Assessment of content and language in light of the new standards: Challenges and opportunities for English language learners. *The Journal of Negro Education, 83*(4), 433–441. doi:10.7709/jnegroeducation.83.4.0433

Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara, CA: University of California Linguistic Minority Research Institute.

Harklau, L. (2000). From the "good kids" to the "worst": Representations of English language learners across educational settings. *TESOL Quarterly*, *34*(1), 35–67. doi:10.2307/3588096

Heatley, E. T. (2002). *How districts and schools utilize data to improve the delivery of instruction and student performance: A case study* (Doctoral dissertation). Retrieved from ProQuest Digital Dissertations Database. (Publication No. AAT 3093769).

Heinrich Heine Universität Düsseldorf (2016). G*Power: Statistical power analyses for windows and mac. Retrieved from http://www.gpower.hhu.de/en.html

Heritage, M., Walqui, A., & Linquanti, R. (2015). *English language learners and the new standards: Developing language, content knowledge, and analytical practices in the classroom.* Cambridge, MA: Harvard Press.

Hernandez, A. (2003). Making content instruction accessible to English language learners. In G. Garcia (Ed.), *English learners: Reaching the highest level of English literacy* (pp. 125–151). Newark, DE: International Reading Association.

Hill, K., Storch, N., & Lynch, B. (1999). A Comparison of IELTS and TOEFL as predictors of academic success. *English Language Testing System Research Reports 2*, 52–63. Retrieved from http://trove.nla.gov.au/work/153133771 ?q&versionId=233552774

Hoover, J. J., & Patton, J. R. (2005). *Curriculum adaptations for students with learning and behavior problems: Differentiating instruction to meet divers needs* (3rd ed.). Austin, TX: Pro-Ed.

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Johnson, E. R. (2009). *Academic language! Academic literacy!* Thousand Oaks, CA: Corwin.

Kenyon, D. M. (2006). *Development and Field Test of ACCESS for ELLs®*. (Technical Report #1). Madison, WI: WIDA Consortium.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*(4), 851– 868. doi:10.1037/0022-0663.100.4.851

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher, 39*(6), 484–486. doi:10.3102 /0013189X10378400

Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade eight. *American Educational Research Journal, 48*(5), 1187–1225. doi:10.3102/0002831211419490

Kim, J. (2011). *Relationships among and between ELL status, demographic characteristics, enrollment history, and school persistence* (CRESST Report No. 810). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Kim, W. G., & Garcia, S. B. (2014). Long-term English language learners' perceptions of their language and academic learning experiences. *Research and Special Education, 35*(5), 300–312. doi:10.1177/0741932514525047

Kinsella, K. (n.d.) Dr. Kate Kinsella on long-term English learners. Retrieved from http://teacher. scholastic.com/products/teach-english-language-learners /english3d.html

Klesmer, H. (1994). Assessment and teacher perceptions of ESL student achievement. *English Quarterly, 26*(3), 8–11. Retrieved from http://www.ncte.org/journals/elq

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York, NY: Pergamon Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: England: Longman.

Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom.* Northumberland, United Kingdom: Bloodaxe Books.

Kumar, P. (2014). Effect of proficiency in English language on academic performance of post graduate management students of Marathwada region (Maharashtra), India. *IOSR Journal of Business and Management, 16*(5), 2319–7668. doi:10.9790/487x-16531016

Lappe. (2000). Taking the mystery out of research: Descriptive correlational design. *Orthopedic Nursing, 19*(2), 81. Retrieved from http://journals.lww.com /orthopaedicnursing/ pages/default.aspx

Latifi, M., Ketabi, S., & Mohammadi, E. (2013). The comprehension hypothesis today: An interview with Stephen Krashen. *Electronic Journal of Foreign Language Teaching, 10*(2), 221–233. Retrieved from http://e-flt.nus.edu.sg/v10n22013 /latifi.pdf

Lehmann, E. L. (2006). *Nonparametrics: Statistical methods based on ranks*. Upper Saddle River, NJ: Springer.

Lindholm-Leary, K., & Genesee, F. (2010). Alternative educational programs for English language learners. In California Department of Education (Eds.), *Improving Education for English Language Learners: Research Based Approaches* (pp. 323–382). Sacramento, CA: California Department of Education Press.

Link, B. G., & Phelan, J. C. (2013). Labeling and stigma. In C. S. Aneshensel, J. C. Phelan, & A. Bierman (Eds.), *Handbook of the sociology of mental health* (pp. 525–541). Dordrecht, Netherlands: Springer Netherlands.

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Englewood Cliffs, NJ: Merrill/Prentice Hall.

Linquanti, R., & Bailey, A. L. (2014). *Reprising the home language survey: Summary of a national working session on policies, practices, and tools for identifying potential English learners*. Washington, DC: Council of Chief State School Officers.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Liu, Y., McGhee, B. D., Kushner, M. I. (2011). The role of bilingual education teachers in preventing inappropriate referrals of ELLs to special education: Implications for response to intervention. *Bilingual Research Journal, 34*(3), 316–333. doi:10.1080/15235882.2011.628608

Lomax, R. (2013). Definitions of quantitative methods of research. Retrieved from http://www.education.com/reference/article/correlational-research/

Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*(2), 177–194. doi:10.1017 /s0272263100004848

Maleki, A., & Zangani, E. (2007). A survey on the relationship between English language proficiency and the academic achievement of Iranian EFL students. *Asian EFL Journal, 9*(1), 86–96. Retrieved from http://70.40.196.162/March_2007_EBook .pdf#page=86

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mathews, G. (2007). Developing the Mountain West Assessment. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 33–45). Davis, CA: University of California.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). New York, NY: Academic Press.

McLaughlin, B. (1990). "Conscious" versus "Unconscious" Learning. *TESOL International Association. 24*(4), 617–634. doi:10.2307/3587111.

Menken, K., & Kleyn, T. (2010). The long-term impact of subtractive schooling in the educational experiences of secondary English language learners. *International Journal of Bilingual Education and Bilingualism, 13*(4), 399–417. doi:10.1080/13670050903370143

Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on "Long Term English Language Learners:" Characteristics and prior schooling experiences of an invisible population. *International Multilingual Research Journal, 6*(2), 121–142. doi:10.1080/19313152.2012.665822

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories* (3rd ed.). New York, NY: Routledge.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan, 96*(7), 64–68. doi:10.1177/0031721715579043

National Clearinghouse for English Language Acquisition and Language Instruction Programs. (2011a). ESL with special needs: National overview. *AccELLerate*, *3*(3), 10–11. Retrieved from http://www.ncela.gwu.edu/files/uploads/17/Accellerate_3_3.pdf

National Clearinghouse for English Language Acquisition and Language Instruction Programs. (2011b). The growing numbers of English Learner Students 1998/99–2008/09. Retrieved from http://www.ncela.gwu.edu/files/uploads/9/growingLEP_0809.pdf

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Author.

National Research Council. (2011). *Allocating federal funds for state programs for English language learners*. Washington, DC: The National Academies Press.

New York City Department of Education, Office of English Language Learners. (2008). *New York City's English language learners: Demographics*. New York, NY: Author.

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

No Child Left Behind Act. (2001). Public Law 107-110. Title III—Language instruction for limited English proficient students. Retrieved from http://www.ed.gov.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417–528. doi:10.1111 /0023-8333.00136

O'Conner, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin, 79*(571), 91–101. doi:10.1037/e687202011-001

Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for California's Long Term English Learners.* Long Beach, CA: Californians Together.

Olsen, L. (2014, March 24). *Meeting the unique needs of long term English language learners*. Retrieved from https://www.nea.org/assets/docs/15420 _LongTermEngLangLearner_final_web_3-24-14.pdf

Oosterhof, A. (2001). *Classroom application of educational measurement*. Upper Saddle River, NJ: Prentice Hall.

Pagano, R. R. (2010). *Understanding statistics in the behavioral sciences* (9th ed.). Belmont, CA: Wadsworth Cengage Learning.

Parker, C. E., Louie, J., & O'Dwyer, L. (2009). *New measures of English language proficiency and their relationship to performance on large-scale content assessments* (Issues & Answers Report, REL 2009–No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center

for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education* (2nd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Race, P. (2009, March). *Designing assessment to improve physical sciences learning: A physical sciences practical guide*. Hull: Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/sites/default/ files/ps0069_designing _assessment_to_improve_physical_sciences_learning_march_2009.pdf

Race, P., Brown, S., & Smith, B. (2005). *500 Tips on assessment* (2nd ed.). London, England: Routledge.

Reeves, J. R. (2006). Secondary teachers' attitudes toward including English language learners in mainstream classrooms. *Journal of Educational Research*, *99(3)*, 131–142. doi:10.3200/joer.99.3.131-143

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. New York, NY: Guilford Press.

Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in an English as a foreign language setting. *Language Testing in Asia*, *3*(12), 12. doi:10.1186/2229-0443-3-12.

Roksa, J., & Potter, D. (2011). Parenting and academic achievement: Intergenerational transmission of educational advantage. *Sociology of Education, 84*(4), 299–321. doi:10.1177/0038040711417013

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*(1), 41–51. doi:org/10.2307/3586011

Sadeghi, B., Kashanian, N. M., Maleki, A., & Haghdoost, A. (2013). English language proficiency as a predictor of academic achievement among medical students in Iran. *Theory and Practice in Language Studies, 3*(12), 2315–2321. doi:10.4304/tpls.3.12.2315-2321

Saffran, J. R., Senghas, A., & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 12874–12875. doi:10.1073/pnas.231498898

Sahragard, R., & Baharloo, A. (2009). Fear of success, imposter phenomenon, academic achievement, and language proficiency among some Iranian EFL learners: A correlational study. *Iranian EFL Journal Quarterly*, *3*(1), 6–35. Retrieved from https://www.iranian-efl-journal.com/447/2009/2014/02/fear-of-success-imposter-phenomenon-academic-achievement-and-language-proficiency-among-some-iranian-efl-learners-a-correlational-study/

Scarcella, R. (2003). *Academic English: A conceptual framework*. Berkeley, CA: University of California Linguistic Minority Research Institute.

Schleppegrell, M. J. (2004). *The language of schooling. A functional linguistic approach.* Mahwah, NJ: Erlbaum.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and

learning: A research review. *Reading and Writing Quarterly, 23*(2), 139–159. doi:10.1080/10573560601158461

Sharkey, J., & Layzer, C. (2000). Whose definition of success? Identifying factors that affect English language learners' access to academic success and resources. *TESOL Quarterly*, *34*(2), 352–368. doi:10.2307/3587961

Shohamy, E., Levine, T., Spolsky, B., Kere-Levy, M., Inbar, O., Shemesh, M. (2002). The academic achievements of immigrant children from the former USSR and Ethiopia. Report (in Hebrew) submitted to the Ministry of Education, Israel. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education, 2nd Edition, Volume 2: Literacy* (pp. 71–83). New York, NY: Springer Science + Business Media.

Simon, M. K. (2011). *Dissertation and scholarly research: Recipes for success*. Seattle, WA: Dissertation Success.

Sivaraman, I., Balushib, A. A., & Rao, D. H. (2014). Understanding Omani students' (University) English language problems. *International Journal of Sciences: Basic and Applied Research (IJSBAR), 13*(1), 28–35. Retrieved from http://gssrr.org /index.php?journal=JournalOfBasicAndApplied&page=article&op=view&path %5B%5D=1708

Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology, 104*(2), 265–285. doi:10.1037/a0025861

Snow, C. E., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49*(4), 1114–1128. doi:10.2307/1128751

Snow, C. E., & Kim, Y. S. (2007). Large problem spaces: The challenges of vocabulary for English language learners. In A. M. R. Wagner & K. K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 123–139). New York, NY: Guilford Press.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In N. Torrance & D. R. Olson (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge, England: Cambridge University Press.

Snow, M., & Brinton, D. (1997). *The content-based classroom: Perspective on integrating language and content.* New York, NY: Longman.

Spolsky, B. (1989). *Conditions for second language learning.* Oxford, England: Oxford University Press.

Sousa, D. (2008). *How the brain learns mathematics*. Thousand Oaks, CA: Corwin Press.

Stella, M. F., Batalova, J., & Fix, M. (2012). *The educational trajectories of English Language Learners in Texas*. Washington, DC: Migration Policy Institute.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Mahwah, NJ: Routledge Academic.

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle Rivers, NJ: Merrill/Prentice Hall.

Suárez-Orozco, C., & Suárez-Orozco, M. M. (2001). *Children of immigration*. Cambridge, MA: Harvard University Press.

Sullivan, A. L. (2011). Disproportionality in special education: Identification and placement of English language learners. *Exceptional Children*, *77*(3), 314–334. doi:10.1177/001440291107700304

Swain, M. (1995). Three functions of output in second language learning. In. G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). New York, NY: Oxford University Press.

Swartz, R. (2001). *Infusing critical and creative thinking into content instruction*. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (pp. 266–274). Alexandria, VA: Association for Supervision and Curriculum Development.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Thomas, R. M. (2005). *High-stakes testing: Coping with collateral damage*. Mahwah, NJ: Erlbaum.

Thomas, W. P., & Collier, V. P. (1997). *Our findings: The "how long" research*. Washington, DC. National Clearinghouse for Bilingual Education.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement: Final report*. Santa Cruz, CA: CREDE

Thompson, K. D. (2015). Questioning the long-term English learner label: How categorization can blind us to students' abilities. *Teachers College Record, 117*(12), 1–50. Retrieved from http://www.tcrecord.org/library/abstract.asp ?contentid=18152

Trochim, W. M. K. (2006, October). Statistical power. Retrieved from http://www.socialresearchmethods.net/kb/power.php

Tutz, S., & Scholz, T. (2003). *Ordinal regression modelling between proportional odds and non-proportional odds*. Paper 304. München, Germany: Institue Für Statistik Sonderforschungsbereich 386. Retrieved from http://epub.ub.uni-muenchen.de/

U.S. Department of Education. (2013). National Center for Education Statistics, Common Core of Data (CCD), "Local education agency universe survey," 2002–03 through 2011–12. Retrieved from http://nces.ed.gov/ccd/pubagency.asp

U.S. Department of Education. (2016). Every student succeeds act. Retrieved from http://www.ed.gov/essa

Valdez-Pierce, L., & O'Malley, J. M. (1992). *Performance and portfolio assessment for language minority students.* Washington, DC: National Clearinghouse for Bilingual Education.

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

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

Wakeman, S. (2013). *Using the ACCESS test as a predictive measure of English learner success on the Biology End-Of-Course-Test in Georgia* (Doctoral dissertation). Retrieved from https://search.proquest.com/openview /f049a0b175de7910412e1823269f918b/1?pqorigsite=gscholar&cbl=18750 &diss=y

Walker, A., Shafer, J., & Liams, M. (2004). Not in my classroom: Teacher attitudes towards English language learners in mainstream classroom. *NABE Journal of Research and Practice*, *2*(1), 130–160. Retrieved from https://www .researchgate.net/publication/237626934_Not_In_My_Classroom_Teacher _Attitudes_Towards_English_Language_Learners_in_the_Mainstream _Classroom

Wertsch, J. V., & Tulviste, P. (1992). L. S. Vygotsky and contemporary developmental psychology. *Developmental Psychology, 28*(4), 548–557. doi:10.1037//0012 -1649.28.4.548

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wilbrink, B. (1997). Assessment in historical perspective. *Studies in Educational Evaluation, 23*(1), 31–48. doi:10.1016/s0191-491x(97)00003-5

Wilson, J., & Komba, S. C. (2012). The link between English language proficiency and academic performance: A pedagogical perspective in Tanzanian secondary schools. *World Journal of English Language, 2*(4). doi:10.5430/wjel.v2n4p1.

Windy Desert School District (WDSD). (2006). *WDSD Secondary School Grading Scale (Grades 6–12).* Retrieved January 08, 2016.

Windy Desert School District (WDSD). (2010). *District improvement and AMAO plan 2010–2013.* Retrieved January 08, 2016.

Windy Desert School District (WDSD). (2011). *English Language Program for student success: Procedural reference manual*. Retrieved January 26, 2016.

Windy Desert School District (WDSD). (2016). *Windy Desert School District master plan for ELL success.* Retrieved February 08, 2016.

Windy Desert School District (WDSD). (2017). WDSD ELL Fact Sheet. Retrieved February 11, 2017.

Wolf, M. K., Kao, J., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses. Practice review* (CRESST Tech. Rep. No. 732). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wong-Fillmore, L., & Snow, C. E. (2000). *What teachers need to know about language*. Washington, DC: ERIC Clearinghouse on Language and Linguistics.

World-Class Instructional Design and Assessment (WIDA) Consortium. (2013). *Annual technical report for ACCESS for ELLs English language proficiency test series 203, 2011-2012 administration: Annual technical report No. 8 volume 1 of 3: Description, validity, and student results*. Madison, WI: Board of Regents of the University of Wisconsin System.

World-Class Instructional Design and Assessment (WIDA) Consortium. (2014a). *Annual technical report for ACCESS for ELLs English language proficiency test series 301, 2012-2013 administration: Annual technical report No. 10 volume 1 of 3: Description, validity, and student results*. Madison, WI: Board of Regents of the University of Wisconsin System.

World-Class Instructional Design and Assessment (WIDA) Consortium. (2014b). Mission and the WIDA story. Retrieved from https://www.wida.us/aboutus /mission.aspx

World-Class Instructional Design and Assessment (WIDA) Consortium. (2015). *Annual technical report for ACCESS for ELLs English language proficiency test series 302, 2013-2014 administration: Annual technical report No. 10 volume 1 of 3: Description, validity, and student results*. Madison, WI: Board of Regents of the University of Wisconsin System.

World-Class Instructional Design and Assessment (WIDA) Consortium. (2016). Consortium members. Retrieved from https://www.wida.us/

Xu, M. (1991). The impact of English-language proficiency on international graduate students' perceived academic difficulty. *Research in Higher Education, 32*(5), 557–570. doi:10.1007/bf00992628

Zafar, M. (2009). Monitoring the "monitor": A critique of Krashen's five hypothesis. *Dhaka University Journal of Linguistics, 2*(4), 139–146. doi:10.3329 /dujl.v2i4.6903

Zhao, H., & Maina, N. (2015). *English language proficiency and progress: Students receiving English for speakers of other languages services from 2012 to 2014*. Rockville, MD: Montgomery County Public Schools, Office of Shared Accountability. Retrieved from http://files.eric.ed.gov/fulltext/ED557668.pdf

Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms, grades 5–12.* San Francisco, CA: Josey-Bass.

Zwiers, J. (2014). *Building academic language: Meeting common core standards across disciplines* (2nd ed.). San Francisco, CA: Jossey-Bass.