Effects of Linguistic Modification Accommodation on High School English Language Learners’ Academic Performance

Semra Beckham
Nova Southeastern University, semra.beckham@gmail.com

This document is a product of extensive research conducted at the Nova Southeastern University Abraham S. Fischler College of Education. For more information on research and degree programs at the NSU Abraham S. Fischler College of Education, please click here.

Follow this and additional works at: http://nsuworks.nova.edu/fse_etd

Part of the Educational Assessment, Evaluation, and Research Commons, and the First and Second Language Acquisition Commons

Share Feedback About This Item

NSUWorks Citation
http://nsuworks.nova.edu/fse_etd/3.

This Dissertation is brought to you by the Abraham S. Fischler College of Education at NSUWorks. It has been accepted for inclusion in Fischler College of Education: Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Effects of Linguistic Modification Accommodation
on High School English Language Learners’ Academic Performance

by
Semra Beckham

An Applied Dissertation Submitted to the
Abraham S. Fischler School of Education
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

Nova Southeastern University
2015
Approval Page

This applied dissertation proposal was submitted by Semra Beckham under the direction of the persons listed below. It was submitted to the Abraham S. Fischler School of Education and approved in partial fulfillment of the requirements for the degree of Doctor of Education at Nova Southeastern University.

_________________________ _______________________
Goli Rezai-Rashti, PhD Date
Committee Chair

_________________________ _______________________
Nibaldo Galleguillos, PhD Date
Committee Member

_________________________ _______________________
Ronald J. Chenail, PhD Date
Interim Dean
Statement of Original Work

I declare the following:

I have read the Code of Student Conduct and Academic Responsibility as described in the Student Handbook of Nova Southeastern University. This applied dissertation represents my original work, except where I have acknowledged the ideas, words, or material of other authors.

Where another author’s ideas have been presented in this applied dissertation, I have acknowledged the author’s ideas by citing them in the required style.

Where another author’s words have been presented in this applied dissertation, I have acknowledged the author’s words by using appropriate quotation devices and citations in the required style.

I have obtained permission from the author or publisher—in accordance with the required guidelines—to include any copyrighted material (e.g., tables, figures, survey instruments, large portions of text) in this applied dissertation manuscript.

________________________________________
Signature

Semra Beckham
Name

April _, 2015
Date
Acknowledgments

This doctoral dissertation is dedicated to my daughter Aubry, my son Ceyhan, my father Mustafa, and my husband Gary. Aubry and Ceyhan; I thank you for your support by showing your pride and faith in my pursuit as I went along this long journey. I am most grateful to my father and my husband who inspired and encouraged me to dream and pursue challenging goals in life and to see the bright side of obstacles to move forward. I miss you. I wish to thank Dr. Goli Rezai-Rashti, my committee chair, for her expertise, guidance, and direction during this dissertation process. I am grateful to have been guided through this intensive and challenging doctoral journey by this exceptional professor. I would also like to give a special thank you to Dr. Nibaldo Galleguillos, my dissertation member, for his guidance, support, and expertise.
Abstract


This applied dissertation was designed to explore the relationship between the language complexity of high school academic assessments and the language proficiency of English Language Learners (ELLs) in their academic achievement levels and to examine what accommodation strategies would be the most effective in reducing performance gaps between ELLs and non-ELLs that are due to language factors. Students designated ELL by the school where the study took place scored significantly lower than non-ELL students in teacher-created content area assessments and state-standardized tests. English for speakers of other languages accommodations, such as extended time in completing tasks and assessments and the use of dictionaries and glossaries, seemed to narrow the gap between ELLs and non-ELLs; however, the effect was not substantial.

Research was conducted to determine whether providing English for speakers of other languages linguistic modification accommodations increased student scores. Two groups of students participated in this research: the control group received the standard test, and the experimental group received the modified test. An original 10th-grade reading comprehension test normed on English-speaking students was administered to the control group and the linguistically modified version of the original test was administered to the experimental group. A comparison of the outcomes was assessed to find out whether there was a significant difference in academic achievement between the two groups. This quantitative study was followed by a qualitative study through student interviews to examine whether there was a relationship between the perceptions of ELLs on the usefulness of the accommodation types and their test scores.

An analysis of the data revealed that students with low English language proficiency may not understand the test questions they are expected to answer. As a result, their test scores may not be an accurate measure of the test item construct, but a measure of their limited English skills.
# Table of Contents

## Chapter 1: Introduction
- Statement of the Problem ................................................................. 1
- Definition of Terms ........................................................................... 7

## Chapter 2: Literature Review
- Theoretical Framework ..................................................................... 10
- Language Proficiency and Academic Achievement ......................... 13
- Time Needed for Proficiency in Academic English ......................... 15
- Language Proficiency Standards and Development of Academic English ......................................................... 17
- Common Core State Standards (CCSS) and its Impact on ELLs .......... 18
- Test Accommodations for ELLs ......................................................... 20
- Purpose of the Study ......................................................................... 45
- Research Questions ........................................................................... 46

## Chapter 3: Methodology
- Participants ....................................................................................... 47
- Instruments ....................................................................................... 50
- Procedures ....................................................................................... 52
- Participant Selection ......................................................................... 54
- Data Analysis ................................................................................... 55
- Limitations ....................................................................................... 56

## Chapter 4: Results
- Quantitative Data Analysis ............................................................... 58
- Qualitative Data Analysis ................................................................. 62
- Conclusion ....................................................................................... 73

## Chapter 5: Discussion
- Summary of the Findings ................................................................. 75
- Interpretation of Findings ................................................................. 77
- Context of Findings ......................................................................... 78
- Implications of Findings ................................................................. 79
- Limitations ....................................................................................... 81
- Future Directions ........................................................................... 83

## References ....................................................................................... 85

## Appendices
- A Original and Linguistically Modified Test Item ................................. 95
- B Interview Questions ....................................................................... 100
- C Linguistic Modification Guide ....................................................... 102
- D Significant Statements in Theme 7: Use of Dictionaries in Tests ...... 105

vi
### Tables

1. Descriptive Statistics of the Control and Experimental Groups ..................61
2. Major Themes Extracted From Interview Transcripts ...............................63
3. Significant Statements in Theme 1: Speaking Spanish at Home and English at School .......................................................................................64
4. Significant Statements in Theme 2: Understanding the Lesson in English ...........................................................................................................65
5. Significant Statements in Theme 3: Understanding Difficult English Words ........................................................................................................67
7. Significant Statements in Theme 5: Understanding the Test Questions....71
8. Significant Statements in Theme 6: Use of Easy to Understand Questions in Tests ........................................................................................72
10. Significant Statements in Theme 9: Differences Between the Two Tests .........................................................................................................74

### Figures

1. Frequency Distribution Histogram of Scores for Test 1 .........................59
2. Frequency Distribution Histogram of Scores for Test 1 .........................60
Chapter 1: Introduction

Statement of the Problem

The nature of the problem addressed by this study was that English language learners (ELLs) were not able to achieve passing scores in academic assessments due to test language complexity. Abedi and Sato (2007) explained that ELLs may spend more time on understanding the language of the test than on answering the questions addressing the content. Abedi and Sato stated that concentrating on the language of the test that is content irrelevant may unnecessarily slow down student performance, and the results of the test may not be valid concerning students’ academic content achievement. Research studies included suggestions that educational assessment of ELLs should not contain language as a factor that influences test outcomes (Abedi, 2006a; Acosta, Rivera, & Shafer Willner, 2008; Young, 2009). Accommodations are intended to remove or reduce the impact of irrelevant and complex language, thereby making assessments accessible and comprehensible. However, research on testing accommodations being used has resulted in mixed outcomes regarding their overall effectiveness with ELLs (Abedi & Sato, 2007; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007).

A study was conducted by Hemphill and Vanneman (2011) in collaboration with the National Center for Education Statistics that covered the period between 1992 and 2009. They found out that the academic achievement gap between ELLs and non-ELLs was 36% in fourth grade, 47% in eighth grade, and 50% in 12th grade (Hemphill & Vanneman, 2011). Additionally, in the state of Florida, ELLs did not meet the adequate yearly progress in reading and mathematics for 4 consecutive years (Florida Department of Education [FLDOE], 2011).

Evidence of the existing trends that have led to the difficulties in meeting
standards may be grounded in the No Child Left Behind (NCLB) Act of 2001. The purpose of the NCLB Act is “to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments.” As a result of the NCLB requirements, there has been a significant change in states’ assessment procedures and decisions in three areas: (a) determining the student population to be assessed, (b) assessing students, and (c) reporting the results of the assessments. One of the most notable changes in states’ testing practices is the mandated inclusion of ELLs and students with disabilities (NCLB, 2001). The measurement of ELLs’ academic proficiency levels through linguistically complex tests increases the challenges for this student population due to limited English language proficiency (Fairbairn & Fox, 2009).

The setting of this study was a suburban public high school in south Florida with a total student population of 3011. There were 380 ELLs in total from 14 different nationalities, according to the 2014 records of the county public school district. In the 2012-2013 academic year, 62% of 10th-grade, non-ELL students scored passing grades on the Florida Comprehensive Assessment Reading Test (FCAT) compared to 26% of ELLs (FLDOE, 2014a). The FLDOE (2014b) reported the overall graduation rate as 79%, compared to 7% of the ELL population.

Research on the assessment and accommodation of ELL students has shown a substantial performance gap between the ELL population and native English-speaking students. This gap may be partly due to the difference in English language skills between the two groups of students (Abedi & Sato, 2007). In order to increase the validity and reliability of assessments administered to ELLs in English, the use of test
accommodations has been widely suggested. Abedi (2006a) analyzed the 73 accommodations that have been used nationally regarding their effectiveness. The Abedi results indicated that only 11 of the 73 accommodations yielded positive results due to their relevance in meeting the educational needs of ELLs. One of the most effective test accommodations was linguistic modification that aimed to remove the unnecessary linguistic complexity that is irrelevant to the construct being measured (Abedi, 2006a).

The topic. According to Abedi and Dietal (2004), linguistically complex, large-scale, criteria-based, standardized tests and content-area, summative assessments administered to ELLs are not reliable, valid, or fair measurement tools of academic knowledge and skills. According to Abedi and Dietal, ELLs cannot be expected to learn content-based knowledge at the same rate as non-ELLs because they have not yet mastered the English language. Teachers’ instructional delivery and assessment questions may be beyond their language skills for full comprehension. Therefore, the assessment and learning conditions need to be addressed. Contrary to a common myth among some educators, Abedi (2006a) found no evidence of ELLs having less ability to learn content knowledge compared with non-ELLs. These findings point to the high possibility of linguistic and cultural biases being the main underlying issues for the performance gap between ELLs and non-ELLs (Abedi, 2006c).

A similar study by Young (2009) confirmed Abedi and Dietal’s (2004) findings. Young stated, “because all assessments measure language proficiency to some degree, when ELLs take an academic content test, the test measures not only content knowledge but also language proficiency” (p. 3). In such cases, language factors are interfering with the struggle of the ELLs to understand the material, which results in distorting the validity of the test to accurately assess content knowledge. As an example, mathematics
word problems test target constructs in mathematics knowledge and skills. However, complex language and unfamiliar words irrelevant to content may prevent ELLs from understanding the questions and from responding accurately, even if they have the knowledge being tested.

Young (2009) identified linguistic difficulty as a construct-irrelevant variance in content assessments for ELLs whose English language proficiency is still developing. The main purpose of content assessments is to measure a student’s academic knowledge or proficiency; therefore, the language of the test items should not create an obstacle for the ELL to show what the learner knows or can do (Young, 2009).

A large number of accommodation recommendations have been suggested to help close the gap between ELLs and non-ELLs in academic assessments. However, a majority of these recommendations do not address language needs, such as extended time, testing ELLs in small groups, better lighting conditions, large print, and frequent breaks during testing, and reading test instructions aloud. These indirect accommodations that are not language-related have been found to be less effective compared to direct linguistic accommodations, such as bilingual dictionaries, glossaries, and translations; and nonlinguistic representations, such as diagrams, charts, and pictures, accompanying test questions. Among the language-related accommodations, linguistic modification has been found to be the most promising. It does not affect the validity of assessments and it has been shown to be more effective in narrowing the performance gap between ELL and non-ELL students (Abedi, Hofstetter, & Lord, 2004; Fairbairn & Fox, 2009; Honigsfeld & Giouroukakis, 2011).

**The research problem.** Adolescent ELLs at the high school where the study took place were unable to achieve the expected competency levels in the linguistically
complex academic tests in English required for high school graduation due to limited English language proficiency. Accommodations in use are not effective in narrowing the achievement gap between native English-speaking students and ELLs. The research tested whether linguistic modification of test items would increase ELLs’ scores on an academic test compared to the scores of ELLs who were not provided with the same accommodation.

**Background and justification.** ELLs are one of the fastest growing populations among kindergarten to Grade 12 students in the United States. According to the National Center for Education Statistics, the number of ELLs in the U.S. public schools increased from 3.7 million in the 2000-2001 school year to over 4.7 million in 2009-2010 (Aud et al., 2012). Overall, student achievement in U.S. schools seems to be increasingly dependent on the academic achievement of ELLs who are expected to represent 40% of the total student population by 2050 (Goldenberg, 2008).

The NCLB Act (2001) prioritizes high-stakes testing, and district and state accountability requiring the inclusion of ELLs in its assessment mandates for school reform. ELLs are expected to make adequate yearly progress, as evidenced by scores on standardized tests of both English language proficiency and subject matter knowledge and skills (NCLB, 2001). However, according to data obtained by the U.S Government Accountability Office, ELLs did not meet language arts and mathematics performance goals in about two thirds of the 48 states from 2005 to 2007 (U.S. Government Accountability Office, 2006).

The American Educational Research Association (2000) position statement on high-stakes testing clearly states, “unless a primary purpose of a test is to evaluate language proficiency, it should not be used with students who cannot understand the
instructions or the language of the test itself” (p. 4). Nevertheless, the NCLB federal
reform act mandates that ELLs are included in standardized tests and state accountability
programs. There seems to be a lack of consideration of the harmful consequences, such as
school accountability reports and dropping out of school (American Educational
Research Association, 2000).

In New York City, which has a large population of ELLs, the ELL graduation rate
was only 25% compared to an overall graduation rate of 69% in 2007, with higher
dropout rates for ELLs compared to non-ELLs (New York State Department of
Education, 2008). This considerable academic achievement gap increased even more
after the state was mandated to include the ELL population in the state’s standardized test
(New York State Department of Education, 2009). The year before the inclusion mandate
took effect the ELL dropout rate was 21% as compared to 16% for non-ELLs. Since the
implementation of the inclusion, the dropout rate for ELLs increased to 29%, yet the
dropout rate for non-ELLs has averaged 17% (New York Department of Education,
2008).

In spite of these bleak statistics, ELLs are still required to be included in high-
stakes accountability programs. In a study conducted in collaboration with Educational
Testing Services, Hakuta (2011) found that the major validity threat related to the
assessment of ELLs seems to originate from factors irrelevant to the knowledge and skills
being measured. Therefore, Hakuta suggested that if English language proficiency is not
the construct of interest, the impact of the students’ level of English language proficiency
must be minimized in order to maximize the validity of the test scores for interpretation.
Hakuta stated that the key to eliminating construct-irrelevant variance was to use clear
and accessible language. However, Hakuta warned against oversimplifying the language
and omitting challenging content vocabulary that might be part of the construct being measured.

**Deficiencies in the evidence.** A majority of the existing studies of ELLs are at-risk students for academic failure focus on improving elementary-level ELL education. However, there is a need for further study on the increased challenges adolescent ELLs face when entering the U.S. educational system at the high school level. The fact that these students have limited time to develop the necessary academic language proficiency essential for academic achievement has not been adequately addressed (Menken, 2009). Furthermore, certain test accommodations may vary in effectiveness depending on grade level; however, this issue has not been widely researched to meet the educational needs of ELLs at the high school level.

**Audience.** This research study is expected to be useful for policymakers, administrators, teachers, and parents. The outcomes discussed in this dissertation study will aid authorities at the federal, state, and district levels: school administrators; and teachers of ELLs to better understand, adjust, and improve the educational needs of this student population. The academic success of ELLs is crucial for the students and the nation.

**Definition of Terms**

The following terms will be used in this applied dissertation and may be unfamiliar to individuals not a part of the field of education.

*English for speakers of other languages* (ESOL) is a term referring to students enrolled in credited English literacy courses designed to improve the English language skills of ELL students through acquisition of communication skills and cultural competencies that enhance ability to read, write, speak, and listen in English (FLDOE,
English language learner (ELL) refers to a person who meets four criteria: first, was not born in the United States and whose native language is other than English; second, was born in the United States, but comes from a home in which a language other than English is most relied upon for communication; third, is an American Indian or Alaskan Native and comes from a home in which a language other than English has had a significant impact on the English language proficiency of that student; and, fourth, has sufficient difficulty speaking, reading, writing, or understanding the English language to deny that student the opportunity to learn successfully in classrooms in which the language of instruction is English (FLDOE, 2009).

English language learner (ELL) classification based on the Individuals With Disability Education Act (IDEA) English Language Proficiency Test, includes five categories:

1. A1: Preproduction phase. Preproduction phase includes non-English speakers or individuals with minimal knowledge of English who demonstrate little understanding, cannot communicate orally, and unable to participate in regular classroom instruction.

2. A2: Early production phase. Limited English speaker demonstrates limited understanding, communicates orally in English with one or two word responses.

3. B1: Speech emergence. Intermediate English speaker communicates orally in English, mostly with simple phrases or sentence responses and makes significant grammatical errors, which interfere with understanding.

4. B2: Intermediate fluency. Intermediate English speaker communicates in English about everyday situations with little difficulty, but lacks academic language terminology and experiences some difficulty in following grade-level, subject-matter
assignments, according to the 2012 records of the county public school district.

5. C1: Advanced English speaker. An advanced English speaker understands and speaks English fairly well, makes occasional grammar errors, and may read and write English with various degrees of proficiency.

*Florida Comprehensive Assessment Test* (FCAT) measures student performance on selected benchmarks in reading, mathematics, writing, and science that are defined by the Sunshine State Standards (FLDOE, 2005).

*Individuals With Disability Education Act* (IDEA) *English Language Proficiency Test* (IPT) is an individually administered measure of speaking and listening proficiency in ESL designed for secondary school students (Stansfield, 1990).

*No Child Left Behind* (NCLB) *Act* is federal legislation that enacts the theories of standards-based education reform. Pursuant to 20 USCS § 6301, the NCLB ensures that all children have access to high-quality, standards-based education.
Chapter 2: Literature Review

This chapter presents a discussion of the theoretical framework within which this study is grounded, a synthesis of the findings concerning the study’s problem area, a discussion of how further research is still needed, and the identification of critical variables in the effectiveness of test accommodations for ELLs. Additionally, important questions to be tested, an indication of shortcomings of prior research, and a critique of the literature as a basis for the methodological decisions in the proposed study are presented.

Theoretical Framework

This study drew on the work of Cummins (1979, 1984) and Messick (1980, 1989) to address the research problem from two different theoretical perspectives: complexity theory of language learning (Cummins, 1984) and meaning in measurement and values in evaluation (Messick, 1975). Cummins’ (1984) complexity theory, described in the seminal work *Wanted: Language Proficiency and Academic Achievement Among Bilingual Students*, is based on the concept of interaction between social and cognitive aspects of second language development, which is a more aggregated and comprehensive approach than some of the earlier methods. The theory includes an explanation of language proficiency as a process developed over time and that each individual’s progress is unique based on a number of variables. Cummins (1984) emphasized the significance of the variability of an individual learner’s language development. Cummins (1979, 1984, 2000) was also a pioneer in distinguishing Basic Interpersonal Conversational Skills (BICS) and Cognitive Academic Language Proficiency (CALP) as two different language proficiencies. BICS refers to the everyday language used in social situations, which is context-embedded (i.e., speakers and listeners can make use of many
cues besides language to communicate). In contrast, CALP is the basis for an individual’s ability to achieve grade-level expectations with academic demands. Cummins (1984) explained that language tasks are more difficult for students when the context is reduced, and extra-linguistic cues are not available. According to the researcher, many children develop native speaker fluency (i.e., BICS) within 1 to 2 years of immersion in the target language; however, it takes between 5 and 7 years for a child to reach native speaker competency in academic language skills. In order to answer the research questions, literature was reviewed relating language proficiency to academic achievement and the distinction between BICS and CALP (Cummins, 1979, 1984, 2000).

This theoretical framework was relevant to the research study as it provided an explanation of reasons why ELLs may appear to be proficient in the language of social interaction, but struggle with the heavy load of academic language with complex structures, academic vocabulary, and complex discourse patterns. It also helped explain why ELLs entering the U.S. education system at the high school level were unable to develop CALP to cope with academic assessments within the timeframe they were enrolled in school.

Furthermore, literature relating to the validity and reliability of academic assessments administered to ELLs was reviewed within the theoretical framework of psychometrician Messick. In a seminal work, Messick’s (1989) description of validity refers to the degree to which interpretations of test scores are supported by empirical evidence and theory, justifying the purpose of the test. In other words, the validity of a test is not rooted in the test itself, but lies in the specific interpretations and uses intended for that test (Messick, 1989). To clarify the distinction between the psychometric adequacy of a test and the appropriateness of its use, Bachman (2004) expanded on these
issues and states that the validation process involves two different parts. According to Bachman, the first part involves “articulating an interpretive argument, which provides the logical framework linking test performance to an intended interpretation and use, [and the second step includes] collecting relevant evidence in support of the intended interpretations and uses” (p. 258).

Messick’s (1989) theoretical framework was directly related to the research problem of this study, specifically related to test validity and fairness for ELLs. When different population groups display significant differences in means, predictors, or both in a particular test, this difference implies that the selection decision of the test is not based on equal consideration of all groups. It may be evidence that the assessment is neither valid nor fair. Therefore, the appropriateness of the interpretations drawn from test scores, such as standardized tests normed on native English-speaking students, but administered to both native English-speaking students and ELLs, does not meet validation or fairness requirements.

Likewise, Gottlieb (2006) emphasized the steps to be taken for validation. In an example directly related to the research problem of this study, Gottlieb stated that the first step in the validation process begins with target construct determination. The second step involves the determination of expected interpretations to be obtained from the test. The last step involves determining the purpose of the test. Based on Gottlieb’s theory, if ELLs are being tested for English language proficiency, the interpretations to be obtained should be the students’ level of proficiency. The purpose of the test might be placement, determination of progress, or making redesignation decisions (Gottlieb, 2006). Drawing on Gottlieb’s theory, when ELLs are tested for specific content knowledge and skills, language factors included in the test construct might confound the test results, resulting in
validity, reliability, and fairness issues. For example, standardized tests of content, such as reading, mathematics, and science, are intended to assess students’ level of proficiency in the specific content area; however, students’ English language skills (construct-irrelevant variance), may add an unnecessary variance and confound their test performance. A test of math which is used for the purpose of measuring mathematics knowledge cannot be assumed to also measure English skills (Young, 2009). This study reviewed literature related to academic assessment of ELLs and synthesized the findings in three ways: (a) language proficiency and academic achievement; (b) English language proficiency standards and academic English; and (c) effectiveness, validity, and reliability of test accommodations for ELLs.

**Language Proficiency and Academic Achievement**

There are a number of research studies that compare student performance on content knowledge and performance on language proficiency to determine whether students who perform at a specific level on the language assessments perform similarly on the content assessments (Abedi, 2001, 2002, 2004b; Abedi, Leon, & Mirocha, 2003; Abedi, Lord, Hoffstetter, & Baker, 2000). The results of these analyses confirmed that there is a strong relationship between the English language proficiency of ELL students and their performance on content assessments.

In studies of differential student performance, native English-speaking students tend to outperform ELLs (Abedi & Lord, 2001; Abedi et al., 2004; August & Shanahan, 2010; Hakuta, 2011). Researchers inferred from this finding that a lack of academic language skills in the English language is a barrier to students’ demonstration of content area knowledge.

Learning academic content in a language other than one’s native language
requires mastery of the second language (Mahon, 2006). The discrepancy between language proficiency of ELLs and the language demands of assessments has been widely addressed. Mahon (2006) explored the relationship between English proficiency and academic performance for a group of ELLs from four elementary schools. In the study, scores from the Woodcock-Munoz Language Survey, which assesses ELL students’ language proficiency, and the state standardized test, Colorado Student Assessment Program, were examined. Findings from descriptive and inferential statistics showed that English proficiency was significantly related to English academic achievement, even for ELLs, who have been in U.S. schools for 3 years or longer.

In Mahon’s (2006) study, ELLs who scored at the proficiency level in Colorado Student Assessment Program subject areas of English language arts, reading, writing, and math, also scored 3.5 to 4 on the language proficiency test of the Woodcock-Munoz Language Survey, which is in the higher intermediate (Level 3) to fluent range (Level 4). These findings enhance educators’ understanding of the linguistic complexity of assessments and their direct impact on ELL students’ achievement gap in comparison to non-ELLS. The evidence of this study further confirms that English proficiency scores are confounded with standardized high-stakes tests, such as the Colorado Student Assessment Program. Obviously, these tests are not accurately measuring the academic achievement of ELLs, due to language limitations. One question that needs to be asked, however, is why ELLs are being included in high-stakes tests in a language they have not yet mastered and why schools are held accountable for the achievement of this student population, the outcomes of which may have serious consequences, both for students and schools. Based on the findings of Mahon’s (2006) study, ELLs should not be included in standardized tests before they reach a certain proficiency score on English proficiency
Educational innovations that are designed to contribute to the success of native English-speaking students may not be effective for ELLs. Once ELLs acquire BICS, they are considered to be at the same language skill level as their non-ELL peers and, therefore, held responsible for attaining the same academic achievement levels as non-ELLs (LaCelle-Peterson & Rivera, 1994). Research on language learning, however, shows that even though ELLs reach basic interpersonal conversational skills in a fairly short time (1 to 2 years), the ability to use English for academic purposes takes approximately 5 to 7 years to develop (Collier, 1992; Cummins & Ada, 1989; Ramirez, Yuen, Ramey, & Billings, 1991).

The equity concept in assessments points to the opportunity for all students to learn the content covered in the curriculum assessments. However, for ELLs, the equity concept has two dimensions: content proficiency and language proficiency. This second dimension of language proficiency raises important equity, validity, and reliability questions. ELLs who have a previous strong educational background may have the content knowledge and skills, but may not be able to perform successfully on academic tasks. Therefore, the assessment procedures in use may not be fair or impartial and may not, therefore, reveal valid results for ELLs (Cummins, 1979, 1984; Fairbairn & Fox, 2009; Gandara & Merino, 1993).

**Time Needed for Proficiency in Academic English**

Collier (1987) conducted a large-scale empirical research study involving 1,548 ELLs to explore the length of time required to become proficient in English for academic purposes. According to Collier, the research questions addressed the variables of “age on arrival, English proficiency level upon arrival, basic literacy and math skills in the native tests.
language upon arrival, and number of years of schooling in English” (p. 617).

The results indicated that ELL students who entered the ESL program between the ages of 8 and 11 had the fastest achievement rate reaching the 50th percentile on national norms within 2 to 5 years (Collier, 1987). However, ELLs who entered the program between the ages of 5 and 7 years were 1 to 3 years behind the performance level compared to the 8- to-year-old entry group, despite the fact that both group members lived in the United States the same amount of time. Those who arrived between the ages of 12 and 15 years experienced the greatest difficulty and were predicted to require as much as 6 to 8 years to reach grade-level norms in academic achievement. The research study included an explanation of why it took a shorter time for 8- to 11-year-old ELLs as compared to 5- to 7-year-old ELLs by noting the advantage of possessing higher level native language skills that help to develop the second language. As this study’s findings may suggest, ELLs who enter the U.S. school system at the secondary school level face the greatest challenges in academic achievement (Collier, 1987).

Another research study was conducted to compare the differential impact of the grade level at which English language proficiency was achieved on ELLs’ academic developmental trajectories. The study’s focus was the cognitive outcomes of a longitudinal sample of first-time kindergarteners through eighth grade (Halle, Hair, Wandner, McNamara, & Chien, 2012). The results of the study indicated that there is a strong relationship between the development trajectories of ELLs and the grade level at which they achieve proficiency in oral English usage. This large-scale study provides evidence that ELLs who are proficient in English at the beginning of kindergarten perform at or above the academic achievement level as their native English-speaking peers and progress at the same or faster rate. However, ELLs, who become proficient in
English by the end of first grade display a gap in reading achievement compared to native English-speakers that prevails and grows over time. Furthermore, ELLs who do not become proficient by the end of first grade stay behind in reading and math skills, both through fifth and eighth grade, compared to their native English-speaking peers (Halle et al., 2012). The findings of this study clearly indicated the significance and impact of the grade level at which English language proficiency is achieved on ELLs’ academic performance. Therefore, the fact that ELLs who enter the American education system at the high school level are expected to achieve the standards normed on native English-speaking students is clearly not a realistic or feasible goal.

**Language Proficiency Standards and Development of Academic English**

Development of standards is a crucial process in determining progress in learning (Abedi, 2006c). Therefore, English language development or proficiency (ELD/P) standards are essential guidelines for the instruction and assessment of the language development of prekindergarten to 12th-grade, ELL students. However, these standards must include the progression of students’ language learning for academic purposes, as well as social language skills (Bailey & Huang, 2011). A review of the literature on the history of standards-based reform demonstrates a lack of progress in this area.

The NCLB (2001) includes mandates that all states must have English development or proficiency standards and assessments must be directly aligned with these standards in accountability for NCLB Title III funding for English Language Development programs. At the beginning stages, each state or a consortium of states was required to create their ELP/D standards. Adopting and adjusting the existing standards from national organizations, such as Teaching English to Speakers of Other Languages was not considered an option. During the substantial task of conceptualizing and creating
standards for language learning, distinguishing the significant construct of academic English as a part of the language learning process was factored into the standards (Bailey & Heritage, 2008).

CALP (Cummins & Ada, 1989) refers to formal academic learning. Academic language acquisition is more than understanding vocabulary. In association with the development of cognitive abilities, the individual is required to learn new concepts through higher order thinking skills, such as comparing, classifying, synthesizing, evaluating, and inferring. These skills are crucial in all academic areas for academic achievement. On the other hand, social language acquisition generally takes place in everyday, meaningful, social contexts and is not cognitively demanding nor does it require specialized language (Cummins, 2008; Echevarria & Graves, 2007).

The process of acquiring academic language skills takes 5 to 7 years. Therefore, setting high expectations for high-stakes tests for ELLs, who begin their education in the United States at the high school level appears to be unrealistic and unfair. Moreover, a majority of ELLs at the high school level are in traditional ESOL programs with a focus on BICS—not CALP. This further confounds and postpones learning the academic language, no matter how hard the students try to cope with content across disciplines at school (Cummins, 2008). This significant objective of including academic English skills as an essential part of the language development of ELLs was not factored into the standards-based education during the initial phases of the ELD/P reform attempt (Bailey & Heritage, 2008).

**Common Core State Standards (CCSS) and its Impact on ELLs**

As the focus on education increases, the new CCSS (2010) takes on increasing importance. The purpose of this education reform movement was to help ensure
consistency in education across all states and to maintain alignment of educational expectations for college and career success. However, the challenge of meeting these standards for ELLs has heightened the risk of underachievement because CCSS places an increased demand on sophisticated language use for all students. In the introduction section of English Language Arts Standards (CCSS, 2010), the following is stated: “students should demonstrate increasing sophistication in all aspects of language use, from vocabulary and syntax to the development and organization of ideas, and they should address increasingly demanding content and sources” (p. 1). However, the guidance provided for states in the development of ELP standards is vague and limited in its scope and aims (Coleman & Goldenberg, 2012).

The framework for English Language Development Standards includes a statement that the framework is not intended to be an outline of how schools should approach teaching ELLs. Each state or consortia of states is responsible for producing its vision of ELL education and ELL curricula using the framework’s articulation of fundamental language practices (Council of Chief State School Officers, 2012). The framework only addresses the links between the CCSS and state ELP standards, leaving an important and critical element of well-defined common ELP/D standards for all ELLs across the nation. Although this reform effort is intended to prepare all students for their postsecondary educational and career pursuits, it gives little specific acknowledgement to the challenges faced by ELLs, particularly on the usage of sophisticated academic language across disciplines (Hakuta, 2011).

With the increased demand for the usage of academic English in instruction and high-stakes assessments, and no clear and firm standards for ELD/P, ELLs are still expected to achieve grade-level linguistic and academic proficiency and attain passing
scores in high-stakes assessments. The imbalance between high expectations in
cademics and vague and limited ELD/P standard guidelines for ELLs creates further
challenges for ELLs and the staff at the schools that serve them. Empirical evidence has
shown the significant impact of English language proficiency on academic achievement.
Under the new reform circumstances, it has become urgent to first develop a sound
framework for the new ELD/P standards to ensure that it captures the substance of
language complexity reflected by the CCSS (Honigsfeld & Giouroukakis, 2011).

The increased academic language demand of the CCSS has a considerable impact
on states as they revise their existing ELD/P standards. This increased demand poses a
critical challenge to the states’ educational decision makers and to test developers. If
language is integral to the CCSS, content standards may need to be assessed for content
and language. Therefore, test developers need to carefully make a distinction between
content-related language (construct relevant) and language that is irrelevant to the focal
construct (content irrelevant) to ensure that the content being measured is not confounded
with construct-irrelevant language. Although the emphasis on the usage of sophisticated
language appears to promise improvement of ELD/P standards for ELLs, there is no
evidence of progress in this area (Hakuta, 2011).

Test Accommodations for ELLs

With the inclusion of ELLs in standardized tests and school accountability,
considerable attention was raised to the instruction and assessment of this student
population. One of the most important concerns in ELL education is the English language
factor, which creates an obstacle for ELLs to show what they know and can do in
academics. In order to reduce the impact of complex language that is construct irrelevant
in large-scale assessments on the academic achievement of ELLs, the use of
accommodations has been widely practiced. Koenig and Bachman (2004) defined accommodation as “any action taken in response to a determination that an individual’s disability or level of English language development requires a departure from established testing protocol” (p. 101).

The strategy of using accommodations was first introduced in 1975 by the IDEA to support students with disabilities during instruction (Center for Public Education, 2012). In the mid-1990s, this practice was extended to assessments of students with disabilities, as well as ELLs. The lack of a sufficient number of research studies on accommodations specific for ELLs led many states to implement the same accommodations allowed for students with disabilities to ELLs, despite the differences between these two groups’ educational needs (Rivera, Collum, Shafer Willner, & Sia, 2006).

The literature included suggestions that the language proficiency of ELLs and high-stakes assessments that demand strong skills in academic language are contradictory concepts that hinder success among the ELLs. The rapidly growing number of ELLs is an era of strong focus on large-scale assessments and, with significantly lower academic achievement levels than native English-speaking students, has drawn increasing attention to accommodations. The validity of test results of ELLs is a widely debated issue based on extensive research on the effectiveness of accommodations.

Nassaji (2003) compared the process of reading comprehension between ELLs and native English-speaking students. English-proficient students have natural language skills and knowledge of academic English to focus on test content without having to first comprehend test item language. However, the study results showed that there are four integral components in second language reading comprehension that ELLs must acquire
before attaining a reasonable processing speed in reading. The four components follow: first, word recognition; second, graphophonics (one of the language cueing systems based upon analyzing letters and phonemes); third, syntactic structures; and, fourth, semantic structures. ELLs often focus on deciphering the test language before they can address the test item content that slows down and confounds their competency in reading comprehension (American Educational Research Association, 2000).

Acosta et al. (2008) described effective testing accommodation features as:

1. They involve changes in testing materials, testing procedures, or the testing situation in order to allow ELLs to participate meaningfully in assessments.
2. They address the unique linguistic and socio-cultural needs of the student without altering the test construct.
3. Accommodated scores are sufficiently equivalent in scale that they can be pooled with unaccommodated scores. (p. 1)

Direct linguistic support and indirect linguistic support are two types of accommodations currently being used. Plain English versions of the test, bilingual dictionaries, glossaries, and clarification or sight translation delivered orally are examples of direct linguistic support. Extended time, frequent breaks during testing, and testing the subgroup separately are indirect accommodations because they do not provide support for the language obstacle (Shafer Willner, Rivera, & Acosta, 2009). The goal of accommodations is to provide a fair opportunity for ELLs to demonstrate what they know and can do, but the fairness concept also applies to native English-speaking students who do not receive such accommodations. If an accommodation is providing an advantage for ELLs over non-ELLs, the validity issue arises (Abedi, 2006a).

According to Abedi (2008), effectiveness, validity, and reliability make up the variables in the complex practical and technical decision-making process of the use of accommodations. Accommodation strategies that greatly reduce the performance gap
between ELLs and non-ELLs are considered effective. The effectiveness of accommodation strategies also refers to the convenience of the implementation of accommodation strategies (feasibility), particularly in large-scale assessments. Validity refers to the provision of accommodations that do not alter the construct of the assessment. Reliability is described as the consistency of a measurement across examinee groups after repeated implementations. Usage of test accommodations has been viewed as the most effective method to meaningfully include ELLs in large-scale academic accountability systems. However, there is no uniformity in the use of accommodations across the nation. Each state has its assessment accommodations matrix for ELLs, often based on feasibility rather than effectiveness and validity. Usage of glossaries, heritage dictionaries, and extra testing time are the most commonly implemented matrix items (Kopriva et al., 2007).

When states decide which accommodations to use for ELLs in academic tests, they must carefully investigate the needs of this student population based on research. Implementation of randomly selected accommodations may not be an effective approach in reducing the academic achievement gap between ELLs and non-ELLs. A recent research study found that ELLs who received accommodations directly addressing their language needs showed higher achievement scores than those who received nonrecommended accommodations that were not aligned with the linguistic needs of ELLs (Kopriva et al., 2007).

**Effectiveness.** For a test accommodation to be considered effective, there has to be evidence that the accommodation increases the performance of ELLs and reduces the performance gap between ELLs and non-ELLs (Abedi, 2004a). Solano-Flores (2008) discussed the effectiveness of accommodations that do not directly address language
because they have been transplanted from the field of special education, such as enhanced lighting conditions, frequent breaks, and large print. The researcher also expanded on the issue of accommodations from the perspective of appropriateness for different ELL groups and cautioned against assuming that all accommodations are appropriate for all ELLs. For example, using side-by-side dual versions of a test may not be effective for ELLs whose reading proficiency in the native language is limited.

To fairly assess this student population, educational researchers recommend the use of several accommodations to help equate testing conditions in association with English language comprehension. However, an accommodation may have an impact on the ability to assess the construct being measured if it provides an unfair advantage to ELLs over those who do not receive the accommodation (Abedi, Courtney, Miricha, Leon, & Goldberg, 2005). An example of this outcome is presented in a research study by Abedi and Gandara (2006). One of the widely used accommodations for testing of ELLs is the usage of bilingual dictionaries; however, the results of this study showed that by accessing the definition of content-related terms, ELLs who are allowed to use dictionaries might have an unfair advantage over those who did not receive the accommodation. Another accommodation that several states currently use at the elementary level is dual-language test booklets. A group of researchers used National Assessment of Educational Progress (NAEP) mathematics test data for 402 Spanish-speaking, eighth-grade students. Due to the fact that the students received instruction of tested content in English, they were unfamiliar with the academic Spanish vocabulary in the Spanish version of the test. Therefore, dual language test booklets did not provide the advantage as predicted. The study results showed that due to the association of instructional language and test language more research studies need to be conducted to
explore which categories of ELLs may benefit most from dual language test booklet accommodation. In agreement with many other researchers, Duncan et al. (2005) also did not find this type of accommodation feasible because there is an average of 460 languages spoken among the ELL population in the United States.

Among the many research studies conducted by Abedi in collaboration with other researchers, one in particular focuses on four issues concerning the use of accommodations for ELLs: effectiveness, validity, differential impact, and feasibility (Abedi et al., 2005). The researchers questioned the four major factors influencing the quality of tests: first, to what extent accommodations helped reduce the performance gap between ELL and non-ELL students (effectiveness); second, whether accommodation strategies had an unfair effect on performance based on group membership (validity); third, whether student background variables played a role in the performance of the students who received the accommodated assessments (differential impact); and, fourth, whether accommodations were easy to implement or use (feasibility).

The study’s setting was 11 school sites and the participants were 611 Grades 4 and 8 students and 24 teachers. There were 317 ELLs and 294 native English speakers or students who had reached native-like proficiency level in English to exit the ELL program. A science test released by the NAEP for each grade level was administered. The test contained original, multiple-choice and open-ended test items created for and field tested on native English-speaking students. The test was administered under four conditions: (a) no accommodation, (b) usage of an English dictionary, (c) usage of a bilingual dictionary, and (d) usage of a linguistically modified version of the test. The linguistic modification involved the reduction of the construct-irrelevant test language without removing the science content construct that was intended to be assessed (Abedi
Students were distributed test materials at random with the student name and the type of accommodation printed on them. The open-ended test items were scored by two teachers, according to NAEP guidelines and scoring rubrics. Inter-rater reliability indices (percentage of exact and within-one-point agreement between scorers) were measured to ensure that the open-ended scoring was objective.

The results of the study showed that some of the accommodation strategies were more effective than others for specific grade levels. As an example, English dictionary accommodation was effective for Grade 4 students, but not for Grade 8 students. The linguistic modification was more effective than either English dictionary use or bilingual dictionary use. These results showed that assessments in higher grades require higher academic English proficiency skills both in vocabulary and discourse knowledge (Abedi et al., 2005). Linguistic complexity includes word frequency and familiarity, word length, and sentence length. Other linguistic features that may cause difficulty for readers include long noun phrases, long question phrases, passive voice constructions, comparative structures, prepositional phrases, sentence and discourse structure, subordinate clauses, conditional clauses, relative clauses, concrete versus abstract or impersonal presentations, and negation (Abedi, Lord, & Plummer, 1997). ELLs who received no accommodations displayed considerably higher numbers of omitted or not reached test items, which confirm research findings that ELLs first attempt to decipher the language of test items to access the content and, therefore, the cognitive load increases, slowing down performance (Abedi et al., 2005).

**Validity.** Koretz (2008) noted, “Few issues in measurement raise such intense emotions as the assessment of students with special needs: those with disabilities or those
with limited proficiency in English” (p. 281). This statement is directly linked to the validity issues in such tests. Validity is an integrated approach that addresses whether the interpretation of a set of test scores can be supported by theoretical justification and empirical evidence (Koretz, 2008). Standards-based content assessments require grade-level, English language proficiency; however, a majority of ELLs are below grade-level language proficiency. Therefore, the scores of ELLs cannot be viewed as valid indicators of their content knowledge (Abedi, 2002, 2006a; Abedi et al., 2004; Young, 2009).

Furthermore, Young (2009) introduced the concept of differential validity with demographic subgroups, which has been a major research topic in assessment since the 1960s. Differential validity is a process used to investigate whether there are differences in validity between different examinee groups. “Because all assessments measure language proficiency to some degree” (p. 123), the differential validity of assessments administered to ELLs in a language they have not yet mastered is highly influenced by the fact that their scores reflect not only content knowledge, but also English language proficiency. If the target construct to be measured is math knowledge, language proficiency that is required to understand the content of test items becomes a second construct, not intended to be assessed, and thereby confounding the interpretation of test results (Young, 2009).

William (2010) expanded on the importance of construct definition and validity in the design of educational assessments for ELLs. The focus of the William research study was the process of the design of assessments. William suggested that the first step in the procedure should be defining the construct to be assessed because construct interpretations are the basis for validity arguments. When the procedure is reversed and construct definition is second to the interpretation, values of the test designer may have
an impact on the construct. According to William, this process is described as “a shift from making the important measurable to making the measurable important” (p. 277).

In other words, when an ELL is assessed on content knowledge, the construct should only assess the target content and not their language skills that are irrelevant to the content being assessed. Therefore, construct definition should precede assessment design for assessments of special needs students just as it is done for the general population (William, 2010).

In agreement with William (2010), Mahoney (2008) pointed to construct validity as the foundational concept for test validation. According to Mahoney, the interpretation and meaningfulness of test scores are directly linked to the validity of the test. Underrepresentation and construct-irrelevant variance are the two factors that can impact the validity of a test. When the test content is too narrow and does not include the vital dimensions of the construct, it is categorized as a threat to the validity of the test. On the other hand, when the test content is too broad, including an excess of information not intended to be assessed, it poses a threat to the validity argument as a construct-irrelevant variance as well. The presence of irrelevant constructs unnecessarily affects the difficulty level of the test for ELLs (Mahoney, 2008). In a study conducted by Abedi (2006a), one of the questions in a math test was “Circle the clump of eggs in the illustration. . . . [A majority of ELL students taking the exam failed to answer this question due to their lack of familiarity with the words] clump [and] illustration. [When the same item was modified to] Draw a circle around the group of eggs in the picture” (p. 47), the accuracy of the students’ responses increased significantly. This particular test item clearly displays construct-irrelevant variance because it tests math knowledge and application, as well as English language proficiency, which is not a part of the target construct. Research
findings illustrate both the discrepancy of interpretation of ELL test scores based on tests administered in English (with a high load of complex linguistic demand), as well as their use for accountability purposes by educational policymakers (Abedi, 2004b, 2006a, 2008; Abedi & Lord, 2001; Solorzano, 2008; Young, 2009).

A similar study was conducted by Mahoney (2008) on the construct validity of test items with a focus on differential item functioning (DIF) in assessments administered to ELLs. Mahoney wrote that a test item is categorized as “DIF [when examinees are from different groups, such as gender or ethnicity, and have the same ability, but have an unequal probability of answering the items accurately. According to Mahoney, a test item is labeled as] non-DIF,” if examinees who have the same ability, have an equal probability of responding to the item correctly, regardless of group membership (p. 16).

Mahoney (2008) investigated whether group membership (e.g., ELL or native English-speaking), differentially affected the achievement on tests by examining the DIF of mathematics test items on the NAEP. The test item functioning analysis results showed that the linguistic complexity of the test items significantly affected the DIF of test items (Mahoney, 2008). Mahoney provided an analogy of using a stopwatch to measure people’s running speed, which runs slower when used for a particular ethnic group of runners. This example illustrates the difficulties of testing ELLs for content knowledge in a language they have not yet mastered.

Among the few studies on the relationship between the language characteristics of standards-based state tests and ELLs’ performance on the tests, a study conducted by Wolf and Leon (2009) stands out due to its unique approaches to investigate the language demands of test items for ELLs. Three states’ standards-based mathematics and science assessments were examined to find out if there was a relationship between the language
complexity of the test items, DIF, and differential bundle functioning (DBF). DBF is a newly developed mechanism for identifying possible sources of DIF against a group of students (Wolf & Leon, 2009). Wolf and Leon stated that looking for sources of DIF in single test items is a long and tedious task and often tends to show inaccurate information. Therefore, DBF has become a more powerful tool for researchers because sources of DIF are more apparent in sets of items that share common, potentially important, characteristics.

This large-scale study involved rating a total of 542 items from 11 assessments at Grades 4, 5, 7, and 8 in three states based on linguistic complexity through the use of a linguistic coding scheme (Wolf & Leon, 2009). General academic vocabulary and the level of linguistic complexity in an item had the strongest association with degrees of DIF, specifically for ELL students with low English language proficiency. The researchers grouped the test items into three bundles (rated from one to three, with one having the least language complexity) and examined the relationship between the varying degrees of language demands and ELL students’ performance. The results indicated that as the complexity of the language increased, each sample displayed a considerable increase in DBF. Bundle 3 test items contained general academic vocabulary that was not content-specific, and ELLs scored approximately 8 points lower per item than non-ELLs of similar ability. Lower English proficiency ELLs scored as low as 16 points below their native English-speaking peers of similar ability.

The unique approach of this study included disentangling linguistic difficulty from content difficulty to find out which factor had an impact on the test scores. The result of the study provides additional evidence that shows the negative impact of linguistic complexity on ELLs’ achievement on content tests. Test developers should be
cautious of test items that contain a high number of general academic words that may lead to unnecessary linguistic complexity, thereby disadvantaging ELL students (Wolf & Leon, 2009).

**Reliability.** Young (2009) defined reliability as, “Equal precision of measurement across examinee groups” (p. 125). The reliability indicator has been investigated to find out if the meanings and interpretations of test performance are the same across groups using Cronbach’s alpha (Cronbach, 1951). Cronbach’s alpha is a numerical coefficient of reliability, ranging from 0 to 1. When alpha is computed for a test, it is based on its reliability relative to other tests measuring the same construct with the same number of test items. Reliability indicators are significant tools to determine whether the same set of test items would receive the same responses if the same questions are readministered to the same sample of examinees. If the test provides stable and consistent responses over repeated administrations of the instrument, it is considered reliable (Abedi, 2002, 2003; Young, 2009; Young, Cho, Cline, & Stone, 2008). Research studies on the reliability indicator include consistently findings that content assessments administered to ELLs exhibit lower reliability values than those of native English speakers.

Abedi (2004b) conducted a study in a state having a large ELL population with a focus on test reliability. The study instrument was the Stanford Achievement Test Ninth Edition, and the target population included students in Grades 2, 7, and 9. The results showed substantially higher differences in the reliability values as grade levels increased. The highest difference in reliability was observed in science and social science tests in Grade 9 (> .2) with higher values reported for the. In comparison, the reliability values at Grade 2 were highest for both the non-ELL and non-ELL groups. These results suggest that the academic domain in higher grades include an increased number of subjects to be
studied. Therefore, ELLs face an increasing demand for academic English language skills and the achievement gap between ELLs and non-ELLs becomes more substantial as they are promoted to the next grade level (Abedi, 2004b).

Research studies included suggestions that the effectiveness, validity, and reliability factors of assessments administered to ELLs must be a priority for policymakers and test developers in order to provide the most appropriate accommodations to reduce the achievement gap between ELLs and non-ELLs. Research evidence clearly included the fact that the most commonly used accommodations, such as extra time, small-group administration, directions read aloud, and use of dictionaries are not the most efficient or valid accommodations for ELLs (Lazarin, 2006). Karantonis (2007) stated, although there is not a single accommodation that is “unequivocally reasonable” (p. 60), one of the few accommodations that has been shown to narrow the gap between ELLs and non-ELLs is the linguistic modification of test items.

Among the many researchers who contributed to literature on effective accommodations for ELLs, Fairbairn and Fox (2009) discussed the challenges and strategies for ensuring that test language is accessible to ELLs who are still acquiring language skills. This research study, in agreement with many others, indicated that linguistic modification of test items is one of the most effective ways to make the test comprehensible and accessible. Based on the results of the study, Fairbairn and Fox suggested that because language complexity is an integral part of the test development process, test developers of high-stakes tests must analyze the language load of test items, tasks, and texts and remove or reduce content-irrelevant language. Another significant point Fairbairn and Fox emphasized is that the linguistic analysis should be conducted at the beginning of the test development process, not at the editing phase. Fairbairn and Fox
also list suggestions for the use of plain language without altering or watering down the content.

Modification of test language has become a widely researched topic since the introduction of NCLB (2001) and the inclusion of ELLs in high-stakes tests. Kopriva (2000) was one of the primary researchers in full support of linguistic modification as an accommodation in assessing ELLs. In agreement with many other researchers, Kopriva listed suggestions to reduce unnecessarily complex and construct-irrelevant language load:

1. Use short, clear sentences or stems with simple sentence/phrase structure.
2. Use consistent paragraph structure.
3. Use present tense and active voice to the extent possible.
4. Minimize rephrasing or rewording ideas.
5. Use pronouns carefully.
6. Use high-frequency words.
7. Avoid or explain colloquialisms and words with more than one meaning. (p. 36)

In addition to the linguistic modification of tests to increase validity and reliability for ELLs, Kopriva (2000) concluded the article by listing other accommodations. These included nontextual representations, such as graphics, charts, and pictures; adjusting the test development process with consideration of cultural differences; increasing research on ELL test development; student feedback on tests; and norming tests to include the different student populations.

Menken (2010) contributed to the debate on the issues of assessment and accountability mandates of the NCLB for ELLs with a focus on the contradiction of
including this subgroup in high-stakes tests in a language in which they have not yet achieved proficiency. The consequences of this long-ignored discrepancy are illustrated in the high-stakes test scores of ELLs and school accountability reports. ELLs score an average of 20 to 50 percentage points below native English-speaking students on state standardized tests, with the result that schools get penalized for failing to meet adequate yearly progress (Abedi & Dietal, 2004).

Menken’s (2010) approach to linguistic complexity of tests is narrowed down to the perspective of “most frequently used words.” (p. 122). Menken conducted a linguistic analysis of the New York State English Regents exam, which is a requirement for high school graduation and for the school evaluation of adequate yearly progress. The study instrument was a 1,200-word reading comprehension passage about a straw-bale building. The first part of the test contained multiple-choice, comprehension questions about the text. The second part of the exam required students to write a persuasive letter to the director of a local agency who was investigating alternative building materials to convince the director to use straw bales as a building material (New York State Education Department, Office of State Assessment, 2009). These tasks made up one quarter of students’ overall scores on the exam.

According to Nation (2006), in order for ELLs to access and comprehend a text in English, 98% of the words in the text must include the most frequently used words in English. However, there is variability in how many words an ELL knows. If an ELL knows 1,000 of the most frequently used words, the text must contain 98% of the words and word families from the 1,000 most frequently used words (Nation, 2006). Menken (2010) compared the words used in the test to the 2,000 most frequently used words in the English language and the most frequent academic word families identified by Nation.
The study results showed that 71.11% of the vocabulary words were from the list of the 1,000 most frequently used words in the English language, 10.68% of the words were from the 1,001 to 2,000 most frequently used words, and 7.09% were academic words that an ELL would likely acquire at the ESOL program exit level. The remaining 11.03% words were low-frequency words, such as rebar and bale, which were key words in the text, critical for the comprehension of the passage (Menken, 2010).

In agreement with previous research, the evidence from this study included suggestions that linguistic complexity is an obstacle for ELLs. Although the NCLB mandates may have drawn increased national attention for effective education of ELLs, ELLs’ inclusion in high-stakes assessments has had a negative impact on student achievement and school accountability systems. ELLs face major educational obstacles that are ever increasing as standards become more demanding and students are assessed by linguistically complex tests in a language they are still in the process of learning (Menken, 2010).

**Language complexity across disciplines.** Another issue affecting academic test achievement scores of ELLs is the level of language demands across disciplines. The language load in math computation test items is considerably less than social studies or language arts test items. Bailey (2005) conducted a review of Grade 11 standardized test items of mathematics, science, and reading comprehension to determine linguistic demands that are construct irrelevant or that reflect unnecessary levels of sophisticated language. Content-specific language, such as mathematical terminology, was excluded from the review. Three evaluative criteria of the study follow: first, site of difficulty in test items (stimulus passage, stem, or response options); second, language domain (vocabulary, syntax, or discourse); and, third, type of linguistic demand (e.g., uncommon
vocabulary, atypical parts of speech, or idiomatic language). The difficulty of the test items was rated from low to high through the use of a Likert-type scale. Approximately 40 to 60 test items from each subject test were analyzed. Bailey reported the reading comprehension section included authentic published texts, the science section contained items that measured knowledge in “using formulas, lists, visual stimuli, and language-rich problems” (p. 81), and the mathematics section had questions that required use of formulas and word problems with a high load of language.

After conducting an initial reading of all test items to determine the range of potential linguistic demands placed on students, Bailey (2005) developed a qualitative coding scheme. The following example of a test item with potential language demands is not taken directly from the instrument used in the study. It is fabricated and used here to illustrate the potential language demand of a test item. Bailey suggested this passage represents the types of item that the researcher set out to analyze:

Mice were randomly assigned to two diet regimens by a biologist working in his lab. Altogether he tended 14 animals. However, he raised five mice with low protein and nine with normal levels of protein. Then, as he fed them, he monitored their health. After just 3 days, five of the mice began to grow sick. The biologist concluded that lack of protein had reduced the immune systems of these mice to a level subject to disease. (p. 87)

After the analysis of this test item, Bailey (2005) found high-level vocabulary and complex demand of syntax and discourse. Bailey explained the vocabulary and syntax demand by pointing out that the meaning of the word “subject [in] lack of protein had reduced the immune systems of these mice to a level subject to disease, [is uncommon and is used to mean] vulnerable to [rather than its more common meaning, the content of a class as in,] Which is your favorite subject at school?” (p. 96) Furthermore, Bailey found the same word having a high syntactic demand because it is used as a verb in this
sentence structure rather than as a noun, which is the more common usage. Another complex syntactic demand in the example is the initial adverbial clause found in this sentence before the main clause (Bailey, 2005): “Then, as he fed them, he monitored their health. . . . [The less complex subject-verb-object order would be.] He monitored their health as he fed them” (p. 96). The adverbial clause at the beginning of the sentence is construct irrelevant, unnecessarily confounding comprehension (Bailey, 2005).

Bailey (2005) expanded on the discourse demand by stating that the use of cohesive ties (e.g., the pronoun “he [to refer to] the biologist [introduced earlier], and the use of logical and temporal connectors (e.g., ] then [and] however”) increase the language processing demands (p. 96). These features require the reader to make meaningful connections between the information presented in a new sentence and information already presented in prior sentences (Bailey, 2005).

The Bailey (2005) study findings showed that 60% of mathematics and 70% of science test items included uncommon words or used words in an atypical manner. In both subject tests, fewer items were identified as complex syntactic structures. The reading comprehension test had the highest percentages of items that presented a higher degree of difficulty in vocabulary and syntax compared to mathematics and science test items (Bailey, 2005). Bailey’s study results on the language demands across disciplines confirm other study findings (Abedi, 2006b; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005).

Another confirmatory empirical study was conducted by Abedi, Leon, and Mirocha (2003) with a focus on language demands across disciplines. The selected disciplines were reading and writing and math problem solving and math computation. Abedi et al. found varying performance gaps between ELLs and non-ELLS depending on
the subject being tested. The performance gap was the highest for reading and writing, lower for science, and the lowest for math problem solving. There was no performance gap between ELLs and non-ELLS in the math computation test (Abedi et al., 2003).

The Abedi, Leon, and Mirocha (2003) study’s findings add to the many other empirical research study results, which included an investigation of the language load of different subjects. All of the study findings share common evidence that ELLs’ academic performance is clearly affected by the level of the language load that accompanies each subject they study. Higher language demands of certain subjects are inaccessible for ELLs and may be the primary reason why this subgroup exhibits underachievement consistently in these academic areas. When the language load is minimal, ELLs are able to show what they know and can do; with minimal language load, the performance gap between ELLs and their native English-speaking peers decreases or, in some cases, disappears (Bailey, 2005; Fairbairn & Fox, 2009; Menken, 2010).

Lemke (2007) wrote that the language of mathematics has been described as a “unified system of meaning-making” (p. 1). This partially universal language consists of natural language, technical terms, and specialized symbolic notations. Although mathematics seems to have a lower language load, even in math tests, language can be an obstacle for ELLs to show what they know and can do. An empirical study included confirmation of the impact of language in mathematics test items on ELL students’ performance (Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006).

The purpose of the study was to investigate to what extent the linguistic features of a mathematics test affected ELLs’ performance when compared to non-ELLS’ performance in Grade 10 (Shaftel et al., 2006). The results showed that the linguistic features of mathematics test items had a moderate-to-large effect on the performance of
ELLs. According to Shaftel et al. (2006), some linguistic features, such as “the use of ambiguous or multiple meaning words, words that are unclear, colloquial, or slang, or that have multiple meanings depending on context for interpretation, prepositions, pronouns, and complex verbs” had the highest impact on student comprehension of the items and tasks (p. 120).

In an earlier and confirmatory research study, Abedi and Lord (2001) used test items from the NAEP. They administered the actual test to the control group, and the modified test to the experimental group. In the modified test, content task and terminology were left intact, but the language was simplified. The study took place in a southern California school, and the sample population included Grade 8 students in low and average math classes. The scores of the experimental group were significantly higher than the control group.

Abedi, Lord, Hofstetter, and Baker (2000) tested 946 Grade 8 students with various accommodations, including linguistic modification of test items, extended time, and usage of a glossary. The Abedi et al. study results showed that only the linguistic modification accommodation narrowed the gap between the ELL and non-ELL students significantly.

Martiniello (2009) examined nonmathematical linguistic complexity (e.g., language that is irrelevant to the mathematical construct being measured) as a source of DIF in math word problems for ELLs and suggested an alternative assessment method. In this large-scale test that involved 68,839 Grade 4 students, Martiniello investigated whether there was a significant difference in test scores between examinees who took the test version with nonmathematical linguistic complexity and the examinees who took the same test with symbolic or visual forms of representation, replacing the complex
language. The findings of the Martiniello study showed that there was a direct negative correlation between nonmathematical lexical and syntactic complexity and student scores. Students who took the test version that included nontextual clues, such as visual displays, graphs, diagrams, and figures, scored considerably higher than the students who took the test version with a high degree of construct-irrelevant language.

In a confirmatory research study, Barton and Neville-Barton (2004) used similar math questions using textual modes of representation (language only), and nontextual, nonlinguistic modalities (visual cues). The results showed that ELLs demonstrated a greater understanding of nonlinguistic or nontextual modes of representation than of textual modes of representation. Another notable result was that both groups of students who were of comparable math skills did not show any comprehension differences in the nontextual version of the test. However, there was a significant difference in the results of the text only math test. Study results showed that the test items with greater linguistic complexity showed positive DIF, favoring non-ELLs over ELLs. On the other hand, items with nontextual representations (abstract images showing connections and relationships between elements), and pictorial representations (concrete images) displayed negative DIF, favoring ELLs over non-ELLs.

**Test language versus instruction language.** Wright and Li (2008) explored the issue of ELLs and their educational needs from a different perspective based on the instructional versus test language used in the classroom. Wright and Li came to the same conclusion as other researchers on the topic of lack of validity in the assessment of ELLs. Although ELLs who have been in the U.S. education system for less than a year are exempt from taking other subject standardized high-stakes tests, the NCLB (2001) mandates require that they take the state standardized math test, even if they have been
enrolled for less than a year.

Wright and Li (2008) provided an analysis of the NCLB policy to include all ELLs, regardless of the time they have been exposed to the English language, through an investigation of fifth-grade, Cambodian students who were new to a Texas middle school. Wright and Li compared the linguistic load of the math work the students completed in class to the Texas Assessment of Knowledge and Skills (TAKS) math test. Educators in Texas offer the test in English and Spanish only; therefore, non-Spanish-speaking ELLs were only able to take the test in English.

After analyzing and comparing the linguistic complexity of classroom assignments to that of TAKS, Wright and Li (2008) argued that the linguistic demands of the Math TAKS tests are not reasonable for newcomer ELLs. The Cambodian students received extra academic support focused on math for 6 months before taking the TAKS test. A wide range of research-based strategies was utilized by the educators to prepare the Cambodian students for the TAKS test at the end of the school year. The strategies implemented during the school year included differentiated instruction with materials that were appropriate and aligned with the students’ language and academic proficiency, extra ESL instruction with an experienced ESL instructor, provision of extra time in the computer lab, ongoing individualized support from a paraprofessional, and primary language support through the use of written Khmer materials. The math teacher worked with the students utilizing a large number of worksheets at lower grade levels and then gradually progressed to upper-grade levels to prepare these students for the upcoming TAKS test. Six months later, all the Cambodian students failed the test. Out of a total of 44 questions, the students got six to eight questions correct on average.

Wright and Li (2008) examined the linguistic complexity of the worksheets used
in instruction and of the TAKS math test items, based on lexical and syntactic levels using Web Vocabulary Profiler, which is a University of Quebec at Montreal-based online research tool developed to identify the difficulty levels of vocabulary and syntax (Cobb & Goldenberg, 2012). The analysis showed that the mathematical lexical density for TAKS math test items was 47.0%, compared to the student math worksheet mathematical lexical density, which was 26.0%. At the syntactic level, the test items revealed even more difficulty in comprehension. As an example, 225 (91.5%) of 246 sentences in the worksheets had basic subject-verb-object sentence structures. However, in the TAKS test, only 66 (55%) of 118 sentences had the basic subject-verb-object sentence structures. In the conclusion of the article, Wright and Li pointed out that despite the fact that the Cambodian ELLs followed a well-organized and implemented regime to learn English and math within 6 months of their arrival in the United States and showed considerable progress both in English language and math, they failed the test. The results of the Wright and Li study showed that the NCLB’s high expectations policy for all students is neither reasonable nor realistic, because language poses an extreme obstacle for ELLs.

Selection of accommodations at the state level. Policymakers and educators have to decide on the selection of accommodations to use by reviewing research-based evidence of their effectiveness and validity. They also must decide which accommodations best meet the needs of various student subgroups and how to implement them (Young, 2009). Of the 73 accommodations currently being offered to states by the U.S. Department of Education, only 11 of them were shown to be highly relevant to the educational needs of ELLs (Abedi, 2006a).

A survey was launched by a group of researchers through interviews to determine
how states decide what accommodations to use for ELLs (Pitoniak, Lutkus, Cahalan-Laitusis, Cook, & Abedi, 2006). State representatives were asked what type of criteria they used for such decisions, what their policies were regarding the inclusion of ELLs in state assessments, and whether research findings were used as criteria in their decisions. The study involved eight states with large numbers of ELLs. The interview results suggested that states generally made their decisions based on recommendations and guidelines provided by the State Department of Education. Only two of the eight states indicated their preference to be language related when making such decisions. When the survey results were analyzed based on states’ decision criteria, it was clear that research findings did not influence the decision-making process. Two other states indicated that the research-based evidence was a priority during the decision process. However, when they were asked about the specific research evidence that they had previously used in their decisions, the state officials were unable to specify them.

Results from experimental studies included a suggestion that accommodations that are language-based or aligned with students’ language needs are more effective and valid for ELLs than those that provide support unrelated to language, such as extended time, testing in a separate room, better lighting conditions, frequent breaks during testing, and oral instructions (Hakuta, 2011). However, the interviews showed that only two of the eight states indicated using accommodations that have direct relevance to ELL students’ language needs (Pitoniak et al., 2006). Research from studies on assessment and accommodation of ELLs can help to identify issues surrounding assessment and accountability systems for ELL students. Findings of individual research studies by Abedi (2002, 2004b, 2005, 2007), as well as those Abedi conducted in collaboration with other researchers, consistently demonstrated that unnecessary linguistic complexity,
irrelevant to the construct being measured, may jeopardize the validity of the assessment and accountability systems. These study results provided suggestions for improving the assessment and accountability systems through various means, one of which is the linguistic modification of assessments (Abedi & Lord, 2001).

According to Abedi (2006a), the goal of providing accommodations for ELLs is to increase the validity and reliability of content assessments to give a fair opportunity to this student population to show what they know and can do. However, accommodations should not give ELLs an advantage over non-ELLs who do not receive accommodations. Many studies have included confirmations that accommodation of linguistic modification for the assessment of ELLs is highly effective and that simplifying construct-irrelevant language is not a threat to score compatibility (Kiplinger, Haug, & Abedi, 2000; Maihoff, 2002; Rivera & Stansfield, 2001). However, there was evidence that some accommodations may affect the validity of assessments, giving an advantage to ELLs over non-ELLs. In one study, there was an increase in both groups’ performance when glossary use plus extra time accommodations were provided to both ELLs and non-ELLs. Considering the fact that glossary use and extended time are the most commonly used accommodations for ELLs, the results of this study raise concerns. Test fairness and validity are issues that need to be addressed for all students, regardless of group membership (Abedi et al., 1998, 2000).

It was becoming extremely difficult to ignore the issues concerning the assessment of academic content for ELLs. Researchers of empirical studies in the area of ELL assessment have demonstrated the ineffectiveness of currently used academic test accommodations. The results from content assessment results display a significant gap between the two student groups and may have serious consequences for students and
schools. It is of utmost importance that test results of ELLs reveal the same meanings and display the same cognitive ability as the results of non-ELLs (Young, 2009). All of these research results clearly suggested that individual studies have contributed to the field. However, educational decision makers should take measures to expand and build a coherent framework by integrating a multitude of research studies on the validity of test comparability. The suggested integration of validity research would improve the interpretation of academic test results for all students (Abedi, 2006c; Fairbairn & Fox, 2009; Young, 2009).

Challenges in educating an increasing number of ELLs have been amplified by educational policies that enforce accountability agendas, standards-based learning outcomes, and standards-driven tests (Fairbairn & Fox, 2009). However, standards-driven testing seems to be designed as one-size-fits-all and does not comprehensively take into account the needs of linguistically and culturally diverse learners.

Approaches in use at the time of this study and reasons why standards are inadequate are shown widely in research studies. Reconsidering and readjusting a time limit of 1 year before ELLs are obliged to take high-stakes tests has often been discussed among researchers. There was evidence that the tests that are being used nationwide are not appropriate or adequate to measure the academic knowledge and skills of ELLs, due to English language limitations (Abedi, 2004a; American Educational Research Association, APA, & NCME, 1999; Fairbairn & Fox, 2009).

**Purpose of the Study**

The research study included an examination of the effect of linguistic modification as a test accommodation in the assessment of the academic achievement of ELLs. The aim of this study was to present a comprehensive view of the theory and
application of the linguistic modification approach to examine its effectiveness as an appropriate accommodation for ELL students. Furthermore, this study addressed academic content assessments normed on native speakers of English and those proficient in academic English and the effectiveness of testing accommodations used for ELLs.

Linguistic modification as an accommodation had never been used for the academic assessment of ELLs at the study site. The assessment accommodations used at the school where and when the study took place were the usage of dictionaries, extended testing time, and testing ELLs in a separate room.

This mixed-methods study was an exploration of the effect of linguistic test accommodation in the assessment of ELLs in a culturally mixed, public high school in south Florida. In this explanatory, experimental, mixed-methods study with a convergence approach, quantitative and qualitative data were collected from 10th-grade, ELL students. Quantitative and qualitative findings were synthesized in a single discussion section.

**Research Questions**

There were three research questions in this study:

1. Does providing linguistic modification accommodation increase ELL students’ performance in a language arts assessment as compared to a standard testing condition?

2. What are the perceptions of ELL students on the effectiveness of test accommodations to improve their test scores?

3. Is there a relationship between the levels of test achievement and perceptions of ELLs on the test accommodations?
Chapter 3: Methodology

An explanatory, mixed-methods, experimental, research method was used in this study to examine the effect of the linguistic modification of assessments as an accommodation on the academic achievement results of ELLs. Quantitative and qualitative results were integrated with research literature to respond to the three research questions. The quantitative data and results provided a general picture of the research problem, and the qualitative data refined, augmented, and explained the general picture in more detail.

An analysis of the quantitative and qualitative data provided a better understanding of the research problem presented in this paper, more so than either type by itself (Creswell, 2008). This study incorporated the use of a control design. Students were divided into two groups: one group of 15 students (experimental) received the linguistically modified version of an English test, while the second group of 15 students (control) took the standard version of the test.

Participants

Quantitative. This study used a parallel sampling strategy. Different samples for the quantitative and qualitative study phases were drawn from the same population (Onwuegbuzie & Collins, 2007). The target population included a total of 30 ELLs who are 15- and 16-year-old, male and female students in the 10th grade. The students were classified as A2 and B1 language proficiency levels, based on the IDEA Language Proficiency Test. The IPT Second Edition is an individually administered test of speaking and listening proficiency in ESL and is designed for secondary students (Stansfield, 1990). It is administered at the initial ELL school enrollment stage and repeated once a year thereafter to monitor ELL language proficiency progress (FLDOE, 2011).
The ELLs at the A1 level are at the preproduction phase in language acquisition. They demonstrate minimal knowledge of English and are unable to participate in regular classroom instruction. At the A2 level, ELLs demonstrate limited understanding and communicate orally in English with two- or three-word responses. At the B1 level, ELLs communicate orally in English, primarily with simple phrases or sentence responses. The B2 proficiency-level ELLs communicate in English about everyday situations with little difficulty, but lack academic language terminology and experience some difficulty in following grade-level subject matter assignments, according to the 2011 records of the study school district. ELLs who reach the C1 level are removed from the ELL program and integrated with mainstream classes. They receive a 1-year, follow-up service to monitor progress. Students with an A1 classification were excluded from the sample as this group of students’ extreme limitation in English language skills would affect the validity and reliability of the study. Students with a B2 classification were also excluded from the sample as this group of students’ distinctively improved English language skills might have affected the validity and reliability of the study. Members of the sample group were from Latin America with varying time of residence and schooling in the United States.

In the state of Florida, students are required to take the FCAT in ninth grade and again in 10th grade. If they fail the test during these 2 years, they are allowed to repeat the test in 11th and 12th grades. It is a graduation requirement that they pass the FCAT by the end of 12th grade. There was a total of 534 10th-grade students at the school where the study was conducted, 47 of whom were ELLs. In the 2013-2014 school year, 40% of the 10th-grade students who were native English speakers scored passing grades on the FCAT Reading test, compared with 6% of ELLs at the same grade level (FLDOE,
None of the participants in the study achieved a passing score on the FCAT in the 2013-2014 school year, according to the school district records.

**Qualitative.** Five ELLs selected through a purposeful sampling process were the target study population (Creswell, 2008). Three ELLs were selected from the experimental group and two from the control group. In purposeful sampling, individuals are intentionally selected to learn and understand the central phenomenon.

The aim of the qualitative sampling approach is to select a representative sample from the population and to generalize the results of the study back to the population (Robson, 2002). According to Patton (2001), the standard for the selection of participants for qualitative studies is based on whether they are “information rich” (p. 169). Qualitative sampling generally involves a small number of participants that enables the researcher to conduct a detailed and in-depth study. Selection of information-rich individuals should be guided by judgment and experience of the quality of information to be collected and learn a determination of the utility of the information. Selection must include criteria for choosing the right individuals, such as those participants who are not hesitant to speak, are articulate, and are likely to share experiences comfortably (Polkinghorne, 2005). Therefore, three participants from the experimental group were selected who were best able to add to the understanding of the phenomenon under study (Gay, Mills, & Airasian, 2008). These three participants’ responses augmented and further explained their perception of the usefulness of the linguistically modified test. Two participants were selected from the control group to gain further insight into the relationship between their perceptions of accommodations and their test scores. This information helped to explain and expand on the quantitative result for the mixed-methods research question.
**Instruments**

There were three instruments in this study: (a) an original 10th-grade reading comprehension test, (b) a linguistically modified version of the original reading comprehension test, and (c) a student interview questionnaire (see Appendix A). The original criterion-referenced reading comprehension test at the 10th-grade level measures an individual’s reading skills in comparison to a criterion, such as identifying a theme, the main idea, compare or contrast, cause or effect, problem or solution skills, identification of literary terms, literary elements, structures, and content vocabulary knowledge. The original test was obtained from *Florida Reading, Grade 10: Standards-Based Instruction* (Henricks, 2010). This FCAT test practice book is used district wide in public high schools in Florida where the school is located. The book includes 10 practice tests in preparation for the state-standardized test. There were 10 items in the original and modified, multiple-choice tests, graded on a scale of 100%; each question was worth 10 points. Students marked their answers on Scantron sheets to be digitally scored.

The second instrument was the linguistically modified version of the original test. The language modification was conducted based on Abedi and Sato’s (2007) report on guidelines to develop a linguistically modified assessment. The researcher modified the language of the original test by simplifying the linguistic domains of lexical and syntactic structures and discourse features to remove the unnecessary complexity of the test language that is irrelevant to content for the linguistic modification guidelines (see Appendix B). This modification helped make the test more comprehensible and accessible for ELLs who are at A2 and B1 levels of English language skills as identified by the IPT Second Edition. Long noun phrases, relative clauses, prepositional phrases, abstract presentations, passive voice, cultural slang vocabulary, and negation add an extra
cognitive load while students attend to decipher the language before they can address the item content. These language demands were identified and replaced accordingly, based on previous research results and recommendations (Abedi, 2006a; Abedi & Soto, 2007).

The researcher followed the steps suggested by Abedi and Soto (2007) during the process of linguistic modification: first, defining the domain and constructs (purpose, assessed domain, assessed constructs, content-related language); second, defining the population (student population, student access needs); and, third, application and evaluation of linguistic modification strategies (categorize content items, apply linguistic modification guidelines and strategies, evaluate the linguistically modified items). During linguistic modification of the test, items reading teachers at the site were consulted for professional feedback on the content rigor and to provide reliability and validity checks.

The third instrument for the study was semistructured student interviews. Interviews took place subsequent to testing. Before starting the interview, the purpose of the study and the interview length was conveyed to the participants. In structured interviews, there is a specified set of questions that will elicit information from the respondents. On the other hand, the goal of unstructured interviews is to find out what the participants have experienced through flexible questions that allow new questions to arise as a result of the participant’s previous response (Gay et al., 2008). The researcher used a combination of these two approaches. The purpose of the semistructured interview is to gather enough information to build a framework of themes to be explored (Drever, 1995). Subsequent to the completion of the modified test, three students from the experimental group and two students from the control group were interviewed. The interview questions were intended to explore the kind of accommodations students found most useful in comprehending the test questions and multiple-choice answers (see Appendix A).
Questions were asked by the researcher in English and teacher assistants translated the question orally into the native language of the students (Spanish). Students chose whether to answer in English or their native language. If the native language was chosen or if the researcher had difficulty understanding the student’s response in English, the translator translated for the researcher. After each question, the researcher used clarifying and elaborating probes to obtain additional information, if necessary (Creswell, 2008). These procedures assisted in establishing validity and reliability. Recorded interview data were transcribed into text and concepts were sorted to create themes.

**Procedures**

The type of research used in this study is an explanatory, mixed-methods, experimental study with an interview approach. In a mixed-methods study, the quantitative and qualitative research methods for collecting, analyzing, and collating data results to answer a research question are used. Mixed-methods research has strengths that neither quantitative nor qualitative only studies possess. It provides more comprehensive evidence that offsets the weaknesses of the quantitative and qualitative research. Descriptive analysis of quantitative data and thematic analysis of qualitative data, together, provided a better understanding of the research problem presented in this paper (Creswell & Plano Clark, 2007).

This type of research consisted of two sequential stages: the collection of quantitative data (test scores) followed by the collection of qualitative data (interview responses) to help explain and expand on the quantitative results. The rationale for this approach was that this study focused on a problem faced by an underrepresented subgroup (ELLs). Augmenting and expanding on the specific quantitative findings (i.e., statistical relationships and differences between groups) through qualitative interviews to
collect different voices from those living the experiences helped to answer the research questions by providing a contextual understanding (Creswell, 2008). In an explanatory, mixed-methods design, the quantitative data are collected first and are given priority over qualitative data (Gay et al., 2008). Therefore, in this study, the quantitative data collection preceded the qualitative data collection and represented a major aspect of data collection (Creswell, 2008).

The experimental and control groups were taught the same reading materials, according to the district instructional focus calendar. The procedures for the study included two phases. In the first phase, all participants received the same 10th-grade, reading comprehension test normed on native speakers of English. The quantitative data were collected and analyzed using a \( t \) test to establish comparability between the experimental and control groups.

In the second quantitative phase of the study, the control group received the original version of the reading comprehension test normed on native English-speaking students and the experimental group received the linguistically modified test version. An independent two-sample \( t \) test was conducted to explore if there was a significant difference between the test scores of the two groups (Creswell, 2008).

The third phase included qualitative data collection from interviews, data analysis, and interpretation to offer more detailed, specific information that cannot be gained solely from the results of the statistical test (Creswell, 2008). After listening to the participants’ answers to the interview questions, taking notes, recording the answers, and transcribing the recordings, data were subjected to analytic induction to find common themes. The qualitative analysis and interpretation were used to help explain and elaborate on the quantitative results. Furthermore, the explanatory, mixed-methods design
provided greater confidence in the generalizability of results by augmenting quantitative test data with qualitative data (Creswell & Plano Clark, 2007).

Data were collected, analyzed, and recorded separately for quantitative (test scores) and qualitative procedures (interview data). The qualitative data were analyzed based on ideas suggested by the interview participants, and themes were determined. The qualitative themes were grouped based on similarity of ideas, quantified, compared, and contrasted with the quantitative findings.

The philosophical assumption of the explanatory design is to begin from postpositivism for the quantitative phase and shift to constructivism for the qualitative phase, lending itself to emergent approaches. Therefore, it was determined whether both sets of results converge, with qualitative results confirming the quantitative results to answer the research questions (Gay et al., 2008).

**Participant Selection**

The unit of analysis for this study was a population of 30 ELLs at the 10th-grade level. ELLs have difficulty in speaking, reading, writing, or understanding the English language, specifically academic English. The majority of ELLs fail to achieve proficiency scores on academic tests. They score an average of 20 to 50 percentage points below native English speakers (Abedi, 2006b). The target population included male and female, Latin America ELLs, Ages 15 and 16. The participants’ English language proficiency levels were A2 and B1 (i.e., beginning and lower intermediate proficiency levels; IPT Second Edition). None of the potential participants scored a passing grade on reading on the Florida Comprehensive Assessment Test in 2014, which is required for high school graduation (FLDOE, 2014b).

To avoid bias, the participants were randomly selected for each group from a list
of student identification numbers from the school database. A consent form was sent to the parents or guardians in English and the native language they speak. Permission from the school principal was also obtained for the study. In addition, the study was submitted to Nova Southeastern University’s Institutional Review Board and the school district’s Institutional Review Board for approval prior to the initiation of any data collection.

**Data Analysis**

The explanatory, mixed-methods, experimental design of the study requires a procedure for collecting, analyzing, and collating the quantitative and qualitative research methods for integrative data analysis in a single study (Creswell, 2008). A $t$ test was conducted on the quantitative data to identify how the control and experimental groups compared on the variable of achievement scores. The qualitative interview data were used for thematic analysis to answer the qualitative research question. An integrated analysis of quantitative and qualitative data was used to answer the mixed-methods research question.

Qualitative data from the interviews were analyzed to determine the perceptions of ELLs on the effectiveness of specific accommodations, which confirmed the quantitative data results. Interviews were audiotaped and transcriptions were analyzed by the researcher inductively (Gay et al., 2008). Each separate idea generated by the participants of the interview was grouped into one of the themes. The themes were converged and reduced based on the similarity of the themes. The analyses explored if there was a correlation between the perceptions of accommodations and test scores to answer the mixed-methods research question. Data were analyzed separately, and then results were interpreted to identify areas of convergence (Onwuegbuzie & Collins, 2007).

The partially structured interview questions, both divergent and convergent,
derived from the research questions provided further and more detailed information to confirm or refute the results of the quantitative instrument. Responses to the open-ended interview questions allowed the participants to respond freely without being forced to select from closed-ended questions, which are sometimes found in questionnaires and surveys (Creswell, 2008). The perceptions of ELLs of the accommodations provided for them aided the researcher to better interpret the results of the two-sample, independent $t$ test and to answer the mixed-methods research question.

In summary, following a descriptive analysis of all participants’ scores, the researcher conducted the inferential analysis of comparing group scores. The student interview results were reviewed, classified, and presented in a table. The researcher developed the explanatory, mixed-methods, experimental research report, including the information about the participants and their assignment, the experimental design, and the intervention and materials, and the control over the extraneous variables and observations.

**Limitations**

The first limitation was related to the need of conducting convenience sampling. This sampling approach was the only possibility for this research, due to the limited population of students eligible for the study. The second limitation was the relatively small size of the sample.

Another limitation in the study was the population. It would have been desirable to have a larger sample size to reach more accurate parameter estimates. In addition, there were certain variables from which the research study would benefit, such as previous academic records, migratory patterns, and interrupted schooling experience of the study participants, which were not available for the entire study target group.
The qualitative data were collected through interviews with five participants using the interview protocol (see Appendix A). This study included three research questions to examine the effects of linguistic modification as a test accommodation, the perceptions of ELL students on test accommodations, and if there is a relationship between the levels of test achievement and the perceptions of ELLs on the test accommodations. This chapter provides the results of the quantitative and qualitative data to assist in answering the research questions.
Chapter 4: Results

Quantitative Data Analysis

Descriptive and inferential statistical analysis was conducted to address Research Question 1: Does providing linguistic modification accommodation increase ELL students’ performance in a language arts assessment as compared to a standard testing condition? Two groups of students participated in this research: a control group ($n = 15$ students) who received the standard, unmodified test, and an experimental group ($n = 15$ students) for whom linguistic modification accommodation was provided. An original reading comprehension test normed on English-speaking students was administered to the control group, and the linguistically modified version of the original test was administered to the experimental group. The aim was to determine if there was a significant difference in academic achievement between the two groups, measured by test scores.

The test scores were originally stored in an Excel spreadsheet, but subsequently imported into SPSS Version 20.0, because SPSS provides more advanced features for statistical analysis (Field, 2009). The first stage was to determine if the data were normally distributed because descriptive parametric statistics (e.g., mean and standard deviation) and inferential parametric statistics (e.g., $t$ tests) assume normally distributed data. The frequency distribution histograms of the test scores (see Figures 1 and 2) approximated bell-shaped curves, reflecting normality, justifying the use of parametric statistics.

The second stage of the analysis was to compute descriptive statistics. The results are presented in Table 1.

The median and mean scores for each test in each group were similar, reflecting
normality. The scores for Test 1 had the same range in the control and experimental groups (minimum = 30, maximum = 60) so the standard deviations were similar ($SD = 9.42$ and $7.94$, respectively). In contrast, the range of scores for Test 2 were wider in the control group (minimum = 30, maximum = 65) than in the experimental group (minimum = 60, maximum = 80) so the standard deviations were dissimilar ($SD = 10.49$ and $5.41$, respectively).

![Figure 1. Frequency distribution histogram of scores for Test 1.](image)

In Test 1, the mean score for the experimental group ($M = 46.67$) was a little higher by 2.33 than the mean score for the control group ($M = 44.33$). However, in Test 2 the mean score for the experimental group ($M = 69.00$) was much higher by 22.0 than the mean score for the control group ($M = 47.00$).

The third stage of the analysis was to determine if the variances of the scores in both groups were equal, using Levene’s test. The results of Levene’s test on the scores
for Test 1, $F(28) = .266, p = .610$, indicated that the variances were equal across the two
groups (because $p > .050$). In contrast, the results of Levene’s test on the scores for Test
2, $F(28) = 4.99, p = .034$, indicated that the variances were not equal in both groups
(because $p < .050$).

![Figure 2. Frequency distribution histogram of scores for Test 2](image)

The fourth stage of the analysis was to determine if there was a significant
difference between the mean test scores of the control and experimental groups. The null
hypothesis was that no significant difference existed between the mean scores. The
results of an independent sample $t$ test on the scores for Test 1 assuming equal variances,
$t(28) = -.730, p = .470$, indicated that the null hypothesis should be retained. There was
no significant difference between the mean scores of the two groups for Test 1 at the
conventional .050 significance level (because $p > .050$). In contrast, the results of an
independent sample $t$ test on the scores for Test 2 assuming unequal variances, $t(21) =
-.7.22, p < .001$, indicated that the null hypothesis should be rejected. There was a
statistically significant difference between the mean scores of the control and experimental groups for Test 2 at the .050 significance level (because \( p < .050 \)).

Table 1

*Descriptive Statistics of the Control and Experimental Groups*

<table>
<thead>
<tr>
<th>Descriptive statistics</th>
<th>Test 1</th>
<th>Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td>15.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>30.00</td>
<td>30.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>60.00</td>
<td>65.00</td>
</tr>
<tr>
<td>Range</td>
<td>30.00</td>
<td>35.00</td>
</tr>
<tr>
<td>( Mdn )</td>
<td>43.57</td>
<td>47.14</td>
</tr>
<tr>
<td>( M )</td>
<td>44.33</td>
<td>47.00</td>
</tr>
<tr>
<td>( SD )</td>
<td>9.42</td>
<td>10.49</td>
</tr>
<tr>
<td>Experimental group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td>15.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>30.00</td>
<td>60.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>60.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Range</td>
<td>30.00</td>
<td>20.00</td>
</tr>
<tr>
<td>( Mdn )</td>
<td>46.67</td>
<td>69.00</td>
</tr>
<tr>
<td>( M )</td>
<td>46.67</td>
<td>69.00</td>
</tr>
<tr>
<td>( SD )</td>
<td>7.94</td>
<td>5.41</td>
</tr>
</tbody>
</table>

The results of the \( t \) tests only determined if the differences between the mean scores were statistically significant, meaning that they deviated from those expected by random chance. Accordingly, the probability (\( p \) value) that the mean difference in the scores was caused by random chance was very small (\( p < .001 \)) for Test 2, but unacceptably high (\( p > .050 \)) for Test 1. Testing for statistical significance did not, however, determine if the results were practically significant, implying that the linguistic modification had a meaningful effect in the context of academic performance (Thompson, 1998; Vacha-Haase, 2001). The final stage of the analysis, therefore was to estimate practical significance, by computing the effect sizes, given by Cohen’s \( d \) (i.e., the difference between two mean scores divided by the pooled standard deviation). Using
Ferguson’s (2009) criteria, the effect size for Test 1 (Cohen’s $d = 0.270$) was small, whereas the effect size for Test 2 (Cohen’s $d = 2.770$) was large.

In conclusion, the results provided statistical evidence to address Question 1: Does providing linguistic modification accommodation increase ELL students’ performance in a language arts assessment as compared to a standard testing condition? The evidence indicated that student performance was not significantly or meaningfully different between the control and experimental groups in Test 1, whereas student performance was significantly and meaningfully increased in the experimental group compared to the control group in Test 2.

**Qualitative Data Analysis**

The purpose of the analysis of the interview transcripts ($n = 5$ students) was to gain insight into the meaning of the students’ experiences, by identifying, through their own words, how they coped with the language complexity of high school academic assessments and to examine what accommodation strategies would be the most effective in reducing performance gaps between ELLs and non-ELLs. Two of the students were in the control group and the three were in the experimental group.

After a preliminary review of the interview transcripts, a template was constructed (see Table 2) consisting of a list of codes to identify the major themes that emerged from the researcher’s interpretation of the transcripts. Nine major themes emerged in the responses through inductive interpretation of the common attributes of the significant statements. The major themes were represented by clusters of significant statements (i.e., phrases or sentences extracted from the transcripts). The major themes in Table 2 were considered to represent the most important meanings and invariant properties of the interview transcripts.
Table 2

Major Themes Extracted From Interview Transcripts

<table>
<thead>
<tr>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Speaking Spanish at home and English at school</td>
</tr>
<tr>
<td>2. Understanding the lesson in English</td>
</tr>
<tr>
<td>3. Understanding difficult English words</td>
</tr>
<tr>
<td>4. Understanding difficult English phrases and sentences</td>
</tr>
<tr>
<td>5. Understanding test questions</td>
</tr>
<tr>
<td>6. Use of easy to understand questions in tests</td>
</tr>
<tr>
<td>7. Use of dictionaries in tests</td>
</tr>
<tr>
<td>8. Use of extra time in tests</td>
</tr>
<tr>
<td>9. Differences between tests</td>
</tr>
</tbody>
</table>

After all the significant statements had been coded according to major themes, the coding was reviewed and refined. This process resulted in the classification of the major themes into subthemes, each indicating a different detail of the major themes. The nine tables included below in this chapter provide the evidence for the classification of the major themes and subthemes, based on the verbatim responses of the participants.

**Theme 1: Speaking Spanish at home and English at school.** The evidence for Theme 1 is presented in Table 3 based on the responses to the question, “You speak Spanish at home and the language used at school is English. How do you feel about this?” The responses of some students follow:

[Two students were confused] Sometimes I get confused. . . . It’s confusing.

I get embarrassed if I make a mistake.

[Two students expressed other feelings] One is the Spanish brain, it is working very good for me; the other one is the English brain. English brain is very small
and I have many difficulties especially in learning my classes. . . . Because my English is not developed I have big problems at school.

Table 3
**Significant Statements in Theme 1: Speaking Spanish at Home and English at School**

<table>
<thead>
<tr>
<th>Name</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>It’s not that hard, sometimes I get confused and start speaking English at home, or speak Spanish at school. My parents don’t understand me when I speak English.</td>
<td>Confused</td>
</tr>
<tr>
<td>B</td>
<td>It’s confusing. I speak Spanish at school more than I speak English at home. I guess my English is not as good as Spanish.</td>
<td>Confused</td>
</tr>
<tr>
<td>C</td>
<td>I speak Spanish most of the time, I don’t speak too much during class because my English is not good and I get embarrassed if I make a mistake. . . . Other students understand the teacher so she thinks everybody understands, I don’t like to interrupt and ask, it’s embarrassing.</td>
<td>Embarrassed</td>
</tr>
<tr>
<td>D</td>
<td>Spanish is my native language and I have no problem at home and at school speaking to my family and friends, but it’s like I have two brains. One is the Spanish brain, it is working very good for me; the other one is the English brain. English brain is very small and I have many difficulties, especially in learning my classes.</td>
<td>Difficulties</td>
</tr>
<tr>
<td>E</td>
<td>I want to be bilingual so I like it. But because my English is not developed I have big problems at school.</td>
<td>Problems</td>
</tr>
</tbody>
</table>

**Theme 2: Understanding the lesson in English.** The evidence for Theme 2 is presented in Table 4, based on the responses to the questions, “When you are listening to the lesson your teacher is teaching, what helps you most in understanding the lesson?”

The most frequent subtheme identified the students’ request for the need to use easy words, reflected by following significant statements:

I want the teacher to tell the lesson with simple words.

It’s easier when they use normal words that we can understand.

Use different words, I mean easy words, for hard words. Some words are strange, never heard them before.

Hope they use words that I know.

[The need for the teachers to repeat what they said was emphasized by] I understand better when teachers repeat what they say; [when] they repeat what
they say and they write notes on the board. I understand very well, [and the endorsement to] Speak slowly and repeat what they say.

Table 4

Significant Statements in Theme 2: Understanding the Lesson in English

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant Statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>I want the teacher to tell the lesson with simple words that we know so we can understand. But they speak very fast and use words that I have no clue what they mean.</td>
<td>Need to use easy words</td>
</tr>
<tr>
<td>B</td>
<td>When teacher uses hard vocabulary words–I don’t understand anything. It’s easier when they use normal words that we can understand</td>
<td>Need to use easy words</td>
</tr>
<tr>
<td>C</td>
<td>Use different words, I mean easy words, for hard words.</td>
<td>Need to use easy words</td>
</tr>
<tr>
<td>D</td>
<td>My English is not good so I listen very carefully, but some words are strange; never heard them before.</td>
<td>Need to use easy words</td>
</tr>
<tr>
<td>B</td>
<td>I hope they use words that I know.</td>
<td>Need to use easy words</td>
</tr>
<tr>
<td>C</td>
<td>I understand better when teachers repeat what they say, go over it again</td>
<td>Need teacher to repeat</td>
</tr>
<tr>
<td>A</td>
<td>Some teachers speak with easy words and they speak slowly and they repeat what they say and they write notes on the board. I understand very well.</td>
<td>Need teacher to repeat</td>
</tr>
<tr>
<td>B</td>
<td>Speak slowly and repeat what they say.</td>
<td>Need teacher to repeat</td>
</tr>
<tr>
<td>A</td>
<td>When teachers ask questions in the class I usually don’t raise my hand because I feel embarrassed if I didn’t understand the question and give the wrong answer.</td>
<td>Need to understand the question</td>
</tr>
<tr>
<td>C</td>
<td>If I miss something when I try to understand, it gets harder to understand the whole lesson. It’s very frustrating sometimes.</td>
<td>Need to understand the question</td>
</tr>
<tr>
<td>C</td>
<td>When I understand the question I don’t have a problem even if the multiple-choice answers have difficult words. I can understand them because I know that knowledge.</td>
<td>Difficult words</td>
</tr>
<tr>
<td>D</td>
<td>It helps me when teachers let my friends, my Spanish speaking friends who also have good English, explain what she is saying, or when she uses words that I know.</td>
<td>Need help from friends</td>
</tr>
</tbody>
</table>

The students found it difficult to answer the teachers’ questions in class, highlighting the need to understand the questions, exemplified by their responses:
I feel embarrassed if I didn’t understand the question and give the wrong answer.

If I miss something when I try to understand, it gets harder to understand the whole lesson.

When I understand the question, I don’t have a problem even if the multiple-choice answers have difficult words.

[A further issue concerning the need for friends to help understand the lesson was raised by one student explaining that] It helps me when teachers let my friends, my Spanish-speaking friends who also have good English, explain what she is saying.

The students were asked four questions about the difficulties they experienced when answering test questions in English: “Do you feel that you know the content of a lesson but cannot answer the questions when you take a test on it? Why? . . . Can you usually show what you know and can do in a test in English? . . . Let’s say you studied very hard for a test. What do you worry about the most before you take a test in English? . . . What do you do when you don’t understand the question in a test?” The answers to these questions were classified into three themes: Theme 3: Understanding difficult English words (see Table 5); Theme 4: Understanding difficult English phrases and sentences (see Table 6); Theme 5: Understanding test questions (see Table 7).

**Theme 3: Understanding difficult English words.** The most frequent subtheme among the significant statements in Theme 3: Understanding difficult English Words was classified as vocabulary (see Table 5). Examples follow:

My vocabulary is not good.

There will be many words that I don’t know.

I can’t answer some questions if I don’t know the vocabulary.

I can’t show much knowledge because my vocabulary is not enough to understand everything and respond.

[One student, however, does not worry about vocabulary tests because] When we
have a vocabulary test, I have no problem because I study the words that are
going to be tested.

Table 5

**Significant Statements in Theme 3: Understanding Difficult English Words**

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>I always worry. I worry because I often don’t get passing grades because my vocabulary is not good. I study vocabulary that we learn in class, but sometimes the vocabulary is not the same that I know.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>B</td>
<td>I worry that my English is not enough to understand the test because there will be many words that I don’t know sometimes.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>C</td>
<td>Nobody can explain the words in a test, I can only use the dictionary and so sometimes I can’t answer some questions if I don’t know the vocabulary.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>D</td>
<td>In algebra, I can show my knowledge because I was good in algebra in my country and there are not too many words to understand. I am good in vocabulary tests because I study. But in other subjects, I can’t show much knowledge because my vocabulary is not enough to understand everything and respond.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>D</td>
<td>If I studied for the test, but have problems selecting the right answers, it’s because of difficult vocabulary that I don’t understand. I end up not answering all the questions right.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>A</td>
<td>When we have a vocabulary test, I have no problem because I study the words that are going to be tested. No surprises.</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>C</td>
<td>Some of the words are very difficult.</td>
<td>Difficult words</td>
</tr>
<tr>
<td>A</td>
<td>I study hard and listen very carefully in the classroom and I think I will get a good grade but when I look at the test and see the big words I lose motivation and feel helpless.</td>
<td>Difficult words</td>
</tr>
<tr>
<td>B</td>
<td>I find out lost points because I didn’t understand the words.</td>
<td>Difficult words</td>
</tr>
<tr>
<td>A</td>
<td>Sometimes there are different meanings, how can I know which one is the right meaning? I know my English knowledge won’t help understand them, and I feel helpless.</td>
<td>Different meanings</td>
</tr>
<tr>
<td>A</td>
<td>There are words that have multiple meanings, how am I supposed to know which one is the correct meaning? In one test, it said that character felt blue-I know blue is a color, but people don’t feel colors.</td>
<td>Different meanings</td>
</tr>
</tbody>
</table>

The need for the students to understand difficult English words was highlighted by statements from the respondents. Examples follow:

Some of the words are very difficult.

When I look at the test and see the big words I lose motivation and feel helpless.
I find out I lost points because I didn’t understand the words.

[The difficulties caused by words with different meanings were emphasized by student responding:] Sometimes there are different meanings, how can I know which one is the right meaning?

There are words that have multiple meanings, how am I supposed to know which one is the correct meaning?

Table 6

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Because I see words or idioms that I don’t know and can’t find it in the dictionary, so I cannot answer the question. I think the solution to my success in tests is to know a lot of difficult vocabulary and understand the test. But I came to America last year and I am learning the easy vocabulary now.</td>
<td>Phrases</td>
</tr>
<tr>
<td>B</td>
<td>I feel frustrated and helpless, I think it’s unfair. For example, in our history test, there was a question that had building boom in it—I could not find the meaning of boom, so I guess that a building was bombed because boom is the sound of a bomb in my language. Of course, it was wrong. Later I found out it means increase of buildings—why can’t they just use that and make it easier for ESOL students to understand?</td>
<td>Phrases</td>
</tr>
<tr>
<td>C</td>
<td>The sentences are in a difficult format—like long sentences. . . . Sometimes there are long sentences with connectors like which, who, that is confusing.</td>
<td>Sentences</td>
</tr>
<tr>
<td>C</td>
<td>It takes so long to figure the words and phrases and long sentences out, and I still don’t know if I figured it out correctly.</td>
<td>Sentences</td>
</tr>
<tr>
<td>A</td>
<td>I am better in math test because the language of math is not so difficult, at least there is no long sentences, but in history or science, there are hard words, long sentences that tire me out to figure out.</td>
<td>Sentences</td>
</tr>
<tr>
<td>A</td>
<td>I still don’t understand the sentence because it’s a long sentence. I try to break it up to make is easier, but I can’t always accomplish doing that either.</td>
<td>Sentences</td>
</tr>
</tbody>
</table>

**Theme 4: Understanding difficult English phrases and sentences.** The need for the students to understand difficult English phrases, including idioms (see Table 6) was highlighted by statements. Some student responses follow:

Because I see words or idioms that I don’t know and can’t find it in the dictionary, so I cannot answer the question.

[Referring to the misinterpretation of the idiom] building boom, [This student suggested that the teachers should] make it easier for ESOL students to understand.
The use of long and confusing sentences was criticized by two students, complaining that Sometimes, there are long sentences with connectors like which, who–that is confusing.

I try to break it up to make it easier but I can’t always accomplish doing that either.

**Theme 5: Understanding the test questions.** The need for the students to understand the test questions was highlighted by complaints from all five participants (see Table 7). Some of the responses of the students’ complaints about not understanding the test questions were reflected by the significant statements:

I knew the answer but could not understand the question.

It is often too difficult to understand everything especially the questions in a test; “I feel helpless sometimes when I don’t understand the test.

Sometimes I don’t understand the question.

[My problem is not understanding the test. . . . It takes a long time to understand the questions; they are too complicated.

I feel blocked when I don’t understand a question, when words are difficult and they use long, complex sentences.

We had an English test, and the text was about Thanksgiving celebration. There were words and phrases I could not find in the dictionary. In my country, we don’t have Thanksgiving, so I don’t know the vocabulary for Thanksgiving and don’t understand the test. . . . If they explained the questions in an easy way, I can answer.

[If one student did not understand the question,] I just skip the question and go to the next question because I don’t want to lose time.

**Theme 6: Use of easy to understand questions in tests.** With reference to test accommodations, such as easy to understand questions, dictionary use, and extended time, the students were asked, “Which one of these is most helpful for you when you take a test? Why?” The significant statements in Table 8 indicated that the use of easy to understand questions was the most popular reply, reflected by the following responses:
Easy to understand questions of course!

Easy questions is the best.

Easy to understand questions would be the best.

So if they ask easy questions— I mean easy for me to understand, it would help me a lot.

Easy language that I understand, especially the questions.

**Theme 7: Use of dictionaries in tests.** None of the students endorsed the use of dictionaries in tests, reflected by the significant statements in Appendix D. The main issue appeared to be the difficulties of finding the right meaning of an English word in a dictionary. The difficulties experienced in the use of dictionaries were exemplified by the following significant statements

Yes I find them, but sometimes later I find out it wasn’t the right meaning.

I am not used to the ones they give us so it becomes more difficult to find the meanings. I worry that I will pick the wrong answer because I can’t find the meaning of words in my dictionary.

Some other students shared their thoughts about dictionary use and the need for familiarity with the resource:

[Concerned about not finding the right meaning in a dictionary,] Sometimes it’s very confusing because there are more than one meaning and I don’t know which one is correct.

If they used an easier word, I would understand it I think.

They allow us to use dictionaries but that doesn’t always help.

I find it sometimes but some words and phrases are not in the dictionary so you have to make a guess.

[The length of time taken to use a dictionary during a test was an issue of concern to three students:] I would lose so much time looking in the dictionary so many times.

I will not have enough time if I search the words in the dictionary.
If you don’t know how to use a dictionary fast, then you lose too much time.

Table 7

Significant Statements in Theme 5: Understanding the Test Questions

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant Statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>After the test, I try to find out what that question was and I feel bad when I find out I picked the wrong answer because I knew the answer, but could not understand the question.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>B</td>
<td>If I understand the text and the questions, I can show my knowledge but it is often too difficult to understand everything, especially the questions in a test.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>C</td>
<td>I am always nervous because I feel helpless sometimes when I don’t understand the test.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>D</td>
<td>I worry that I won’t get a good grade because sometimes even when I study for the exam I get bad grades because my English knowledge is not very good and sometimes I don’t understand the question.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>B</td>
<td>My problem is not understanding the test.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>C</td>
<td>No, because it takes a long time to understand the questions; they are too complicated. I usually read the same question many times and give up.</td>
<td>Cannot understand the questions</td>
</tr>
<tr>
<td>C</td>
<td>I feel blocked when I don’t understand a question, when words are difficult, and they use long, complex sentences. It feels like there is a wall around my brain and I can’t work on the test. I know I will pick the right answer if I can understand the question.</td>
<td>Cannot understand the question</td>
</tr>
<tr>
<td>B</td>
<td>For example, we had an English test, and the text was about Thanksgiving celebration. There were words and phrases I could not find in the dictionary. In my country, we don’t have Thanksgiving, so I don’t know the vocabulary for Thanksgiving and don’t understand the test.</td>
<td>Cannot understand the question</td>
</tr>
<tr>
<td>B</td>
<td>If they explained the questions in an easy way, I can answer.</td>
<td>Explain the question in an easy way</td>
</tr>
<tr>
<td>D</td>
<td>I just skip the question and go to the next question because I don’t want to lose time.</td>
<td>Skip the question</td>
</tr>
</tbody>
</table>

Theme 7: Use of extra time in tests. The five students did not endorse the use of extra time, as exemplified by the significant statements in Table 9. Some sample responses follow:

Extra time doesn’t help.
Extra time is useless if you can’t understand the questions.

I don’t think extra time helps me.

I don’t think extra time is useful. . . . It just gives you more time to feel frustrated.

Extra time makes me tired and confused when I go over the questions I can’t understand.

Table 8

Significant Statements in Theme 6: Use of Easy to Understand Questions in Tests

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Easy to understand questions of course!</td>
<td>Easy to understand questions</td>
</tr>
<tr>
<td>B</td>
<td>Easy questions is the best. Because that’s where I stop and spend a lot of time trying</td>
<td>Easy to understand questions</td>
</tr>
<tr>
<td>C</td>
<td>Easy to understand questions. Because if I understand what they are asking I don’t</td>
<td>Easy to understand questions</td>
</tr>
<tr>
<td></td>
<td>need extra time or dictionary. . . . Easy to understand questions would be the best</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>If I understand what the question means I can find the answer. So if they ask easy</td>
<td>Easy to understand questions</td>
</tr>
<tr>
<td></td>
<td>questions–I mean easy for me to understand, it would help me a lot.</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Easy language that I understand, especially the questions.</td>
<td>Easy to understand questions</td>
</tr>
</tbody>
</table>

Theme 9: Differences between the two tests. The significant statements in Theme 9 are presented in Table 10. Responses are from the experimental and control groups:

[Control group student] Each one was the same, not too difficult not too easy.

I think both were not easy.

[Experimental group, however, suggested that the second test was easier to understand:] The second one was easy, I only looked at the dictionary a few times, and I finished early.

The second text was hard too but the questions were easy to understand so it was easy to select the correct answer.

I think I answered more questions correct in the second one because when I read the questions it was easy to understand where I have to look in the text to find out
the answer.

Table 9

Significant Statements in Theme 8: Use of Extra Time in Tests

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Extra time doesn’t help if you don’t understand vocabulary.</td>
<td>Extra time does not help</td>
</tr>
<tr>
<td>B</td>
<td>We get extra time sometimes, but it doesn’t help. . . . Extra time is not very useful, I sit there going over the questions I could not answer until the last minute, but it doesn’t really help get a good grade.</td>
<td>Extra time is not useful</td>
</tr>
<tr>
<td>B</td>
<td>I use the extra time for tests, but I still cannot get good grades. Extra time is useless if you can’t understand the questions. Extra time does not unlock the English vocabulary section of my brain; if it does, there is not much there anyway.</td>
<td>Extra time is not useful</td>
</tr>
<tr>
<td>C</td>
<td>I don’t think extra time helps me unless it is math or algebra or something like that and it’s not the vocabulary that is difficult, it’s the answers I have to figure out in math language.</td>
<td>Extra time does not help</td>
</tr>
<tr>
<td>D</td>
<td>I don’t think extra time is useful . . . it just gives you more time to feel frustrated. When I use extra time, I don’t think it helps me get a better grade.</td>
<td>Extra time is not useful</td>
</tr>
<tr>
<td>E</td>
<td>Extra time makes me tired and confused when I go over the questions I can’t understand.</td>
<td>Extra time makes me tired and confused</td>
</tr>
</tbody>
</table>

Conclusion

The results provided statistical evidence to address Question 1: Does providing linguistic modification accommodation increase ELL students’ performance in a language arts assessment as compared to a standard testing condition? There was no significant difference in student performance between the control and experimental groups in Test 1 (i.e., an original reading comprehension test normed on 11th-grade, native English-speaking students) implying that the English language proficiency of the two groups was equivalent. There was, however, a significantly and meaningfully increase in the scores for Test 2 among the experimental group compared to the control group, implying that the linguistically modified version of the original test administered the experimental group had a beneficial impact on the test scores.
Table 10

**Significant Statements in Theme 9: Differences Between the Two Tests**

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant Statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>The first one had many difficult words; I looked in the dictionary many times. I don’t think I got a good score. It was hard. I even used the extra time. The text was hard and the questions were hard.</td>
<td>First test had difficult words</td>
</tr>
<tr>
<td>B</td>
<td>The first one was difficult maybe because it was about science, I don’t know much about science, I am not good at it. But the vocabulary was very difficult too.</td>
<td>First test had difficult words</td>
</tr>
<tr>
<td>A</td>
<td>The second one was easy, I only looked at the dictionary a few times, and I finished early.</td>
<td>Second test was easier to understand</td>
</tr>
<tr>
<td>B</td>
<td>The second text was hard too but the questions were easy to understand so it was easy to select the correct answer.</td>
<td>Second test was easier to understand</td>
</tr>
<tr>
<td>C</td>
<td>Both were difficult. But I think I answered more questions correct in the second one because when I read the questions it was easy to understand where I have to look in the text to find out the answer. Also the questions were not made up of long sentences which is easier to understand</td>
<td>Second test was easier to understand</td>
</tr>
<tr>
<td>D</td>
<td>Each one was the same, not too difficult not too easy. I understood the first one better because I like science.</td>
<td>Each test was the same</td>
</tr>
<tr>
<td>E</td>
<td>I think both were not easy.</td>
<td>Both were not easy</td>
</tr>
</tbody>
</table>

The results of the qualitative analysis of the interview transcripts corroborated the results of the quantitative analysis of the test scores. The students in the experimental group suggested that the second test was easier to understand, implying that the modified version of the original test was effective in improving their understanding. The students complained that their lack of understanding of English (particularly with respect to difficult words, phrases, and sentences, and complicated questions) was a major barrier and, therefore, the modified test appeared to break down this barrier.
Chapter 5: Discussion

This chapter presents a summary of the study, along with interpretations of findings, the context of findings, implications of findings, limitations of the study, and suggestions for future research. This mixed-methods study was conducted at a suburban high school to examine the effect of linguistic modification as a test accommodation in the assessment of academic achievement of ELLs at the high school level. The sample population was 11th-grade ELLs with lower intermediate and intermediate English language proficiency skills based on the IDEA Language Proficiency Test (Schrank, Fletcher, & Alvarado, 1996). There were three instruments for the quantitative part of the study: an original reading comprehension test normed on 11th-grade, native English-speaking students to find out the comparability of the English language skills of the control and experimental groups. The second and third instruments were an original reading comprehension test normed on 11th-grade, native English-speaking students and a linguistically modified version of the same test to find out if linguistic modification as an accommodation was effective in increasing ELL students’ scores. The fourth instrument was an interview protocol used for the qualitative part of the study to find out the perceptions of ELLs about the accommodations they receive for testing.

The explanatory, mixed-methods, experimental design of the study required a procedure of collecting, analyzing, and collating the quantitative and qualitative research methods for integrative data analysis in a single study (Creswell, 2008). The quantitative data provided an answer to the first research question to identify how the control and experimental groups compared on the variable of achievement scores. The qualitative data were used for thematic analysis to answer the second qualitative research question about the perception of ELLs on the effectiveness of test accommodations. An integrated
analysis of quantitative and qualitative data was used to answer the third research question to find out the relationship between the levels of test achievement and perception of ELLs on the effectiveness of test accommodations.

At the study site, 62% of 10th-grade, non-ELL students scored passing grades on the FCAT compared to 26% of ELLs in the 2012-2013 academic year (FLDOE, 2014a). In the same year, the FLDOE (2014b) reported the overall graduation rate as 79%, compared to 7% among the ELL population. The researcher wanted to investigate if the level of achievement would improve among ELLs using linguistic modification as a test accommodation. A review of the literature indicated that linguistic modification accommodation was not commonly used nationwide at the study site as an accommodation despite the fact that empirical research showed positive results (Abedi et al., 2004; Fairbairn & Fox, 2009; Honigsfeld & Giouroukakis, 2011)

Summary of the Findings

Findings for Research Question 1. The results provided statistical evidence to address Research Question 1: Does providing linguistic modification accommodation increase ELL students’ performance in reading comprehension assessment as compared to a standard testing condition? The evidence obtained from Test 2 indicated that student performance was significantly and meaningfully increased in the experimental group who received a linguistically modified version of the test compared to the control group who received the original test normed on 11th-grade, native English-speaking students.

Findings for Research Question 2. The second question was: What are the perceptions of ELL students on the effectiveness of test accommodations to improve their test scores? According to the interviewed participants, extra time and use of dictionaries were not adequately effective as test accommodations. Although it provided the needed
time to look up unfamiliar vocabulary in the dictionary, difficulties in finding the
appropriate meaning of the word in the case of multiple meanings, or not finding the
word or phrase added to the frustration of taking a test and finishing on time.

Based on the five participants’ interview responses, the questions with simple
language played a significant role in higher test achievement scores. They also
commented that vocabulary that is explicitly taught in the classroom did not usually
appear in tests. This was an unexpected research finding, which is extensively discussed
in an empirical research study (Stevens, Butler, & Castellon-Wellington, 2001). This
study included an investigation of the relationship between content vocabulary taught in
class and the vocabulary used in standardized achievement tests. The study results
indicated that the correspondence between curriculum vocabulary and test vocabulary is
limited.

**Findings for Research Question 3.** The third research question was “Is there a
relationship between the levels of test achievement and perceptions of ELLs on the test
accommodations? The three interview participants from the experimental group who took
the linguistically modified test indicated that the second test was easy because they
understood the questions well and did not spend much time using the dictionaries. These
participants achieved higher scores in the modified test compared to the control group
who took the original test. In addition, the two interviewed participants who did not take
the linguistically modified test indicated that they would prefer test questions to be in
simple language over dictionaries and extra time.

**Interpretation of Findings**

There was a significant score gain for the experimental group who took the
modified test. The results of this study indicated that students with low English language
proficiency may not understand the test questions they are expected to answer. As a result, their test scores may not be an accurate measure of the test item construct, but a measure of their limited English skills. A possible explanation for the results of the quantitative data in this study may be that linguistic modification of the test questions improves the accessibility of the question and increases the validity of items measured in reading comprehension (Abedi, 2006a).

**Context of Findings**

The findings of the quantitative part of the study results are in agreement with the literature review of this study. One of the problems ELLs face in education is the ability to access test items that support comprehension of the academic construct being measured (Sato, Rabinowitz, Gallagher, & Huang, 2010). Due to their limited English proficiency, access to test items is constrained when test language includes unnecessary and irrelevant words or phrases, unfamiliar sociocultural contexts, or culture-specific references to measure the tested skill (Abedi, 2008). When meaningful engagement with the test constructs is limited, construct-irrelevant factors may interfere with the results that accurately measure students’ level of achievement (Hemphill & Vanneman, 2011). Strategies to facilitate student access to linguistically challenging tests should address the linguistic challenges that ELLs face so that they can demonstrate what they know and can do. Another concern for educators is that the interpretations of comparability of the test scores of ELL students with those of native English-speaking students may not be valid due to lack of access to test items by the ELL subgroup (Abedi, 2006a; Sato, Moughamian, et al., 2010; Young, 2009). There was empirical research showing that during testing ELLs may focus on aspects of the language of the test rather than selection of a response based on the targeted content, which leads to lower academic success rate
compared to non-ELLs (Nassaji, 2003). In this research study, the significant score gain of the experimental group indicated that this group of students were able to direct their cognitive resources to select an answer choice, rather than configuring the meaning of the test question.

**Implications of Findings**

This study drew on the work of Cummins (1979, 1984) and Messick (1980, 1989), while addressing the research problem from two different theoretical perspectives: first, complexity theory of language learning (Cummins, 1984); and, second, meaning in measurement and values in evaluation (Messick, 1975). Literature relating to the validity and reliability of academic assessments administered to ELLs were reviewed within the theoretical framework of psychometrician Messick (1989). In a seminal work, Messick’s (1989) description of validity refers to the degree to which interpretations of test scores are supported by empirical evidence and theory, justifying the purpose of the test. In other words, the validity of a test is not rooted in the test itself, but lies in the specific interpretations and uses intended by that test (Messick, 1989). Messick’s (1989) theoretical framework was directly related to the research problem of this study, specifically related to test validity and fairness for ELLs. When different population groups display significant differences in means, predictors, or both in a particular test, this difference implies that the selection decision of the test is not based on equal consideration of all groups. This is evidence that the assessment is neither valid nor fair. Therefore, the appropriateness of the interpretations drawn from test scores, such as standardized tests normed on native English-speaking students, but administered to both native English-speaking students and ELLs, does not meet validation or fairness requirements.
Aguirre-Munoz (2000) explored the use of Spanish-only version and a linguistically modified English version of a test as test accommodations administered to native Spanish-speaking ELLs. Aguirre-Munoz found that there is a strong relationship between the level of English language proficiency of the ELLs and the type of accommodation that best meets the educational needs. ELLs with no or very little English proficiency benefited most from the Spanish-only version while students with intermediate proficiency in English benefited more from the linguistically modified English version of the test. This shows that the implications of this study’s findings are consistent with current research and practice areas used at the time of this study. The participants of the study had lower-intermediate to intermediate level of English proficiency, and the experimental group’s scores were significantly higher than those of the control group in Test 2.

Furthermore, in a study (Abedi et al., 1997), the results suggested that the effectiveness of accommodations may vary across grade levels. Linguistic modification accommodation was shown to be more effective than other accommodations for eighth-grade ELLs, but not for fourth-grade ELLs, probably due to the fact that tests in higher grades may have complex vocabulary and complex discourse. The results of the study are in agreement with empirical research studies completed in the area of accommodation practice that are in use at the time of this study.

The status of the accommodations has changed over time as more research evidence became available. In the past, ELLs were excluded from state standardized tests altogether. Their inclusion started with the NCLB Act of 2001 providing accommodations designed for exceptional student education, which is still in practice. There has been a growing attention to the issues of accommodations by the researchers
and policymakers. Researchers carefully examined the accommodations being used and have provided research-based accommodations. However, in practice, the policymakers’ priority was mainly the effectiveness of accommodations, specifically effect sizes when making decisions on which accommodations to use. The issue of validity did not receive as much consideration. As an example, the usage of a commercial dictionary is a commonly implemented accommodation nationwide. However, studies documented that this accommodation improves the performance of ELL students and non-ELLs. The goal of accommodation is to make an assessment more accessible for English language learners to produce results that are valid for these students. The intent is not to give them an unfair advantage over those who are not receiving that accommodation. This study adds to the existing research studies, which consider both effectiveness and validity of accommodations with a focus on issues of measurement.

**Limitations**

Preimplementation of the study, one of the expected potential limitations related to the design of the study was related to the need to conduct convenience sampling. The convenience sampling is often confounded by bias due to the fact that a convenience sample can lead to underrepresentation or overrepresentation (Creswell, 2008). This sampling approach was the only possibility for this research, due to the limited population of students eligible for the study. Furthermore, regarding external validity issues, in convenience sampling because the sample is not chosen randomly, it is unlikely to represent the target population and, therefore, impairs the researcher’s ability to make generalizations.

The second limitation was the relatively small size of the sample population. One of the disadvantages of this limitation is variability. Variability is determined by standard
deviation of the population; the smaller the sample, the larger the standard deviation and so the less accurate the results. In this study, it would have been desirable to have a larger sample size to reach more accurate parameter estimates.

In addition, the study’s internal validity was threatened by the fact that there were certain variables from which the research study would benefit, such as the previous academic records, migratory patterns, and interrupted schooling experience of the study participants, which were not available for the entire study target group. Therefore, without differential impact knowledge, the outcome may not accurately reflect whether linguistic accommodation was the only cause for the raised achievement levels of the experimental group or other possible causes, such as individual student academic background, mastery of native language, the length of time the student has resided or schooled in the United States, the language spoken at home, educational status of parents or guardian, and socioeconomic background.

The researcher followed the guidelines for linguistic modification suggested by Abedi et al. (1997). Linguistic modification is a theory and research-based process for changing the language of the test items for clarification without altering the target construct being assessed (Abedi 2008; Abedi et al. 2005). Linguistic modification is intended to increase student access to tested content by removing unnecessary and test-construct irrelevant language that could place ELLs at a disadvantage. However, accommodations should not give an unfair advantage to ELLs over non-ELLs, which may have an effect on the reliability and validity of the test. This was an additional limitation in the study because the study did not include non-ELLs to find out if the accommodation was also effective for non-ELLs with higher achievement levels compared to the original test.
Future Directions

The effectiveness of current test accommodation for making assessments accessible, equitable, and valid for ELLs are vague due to the fact that little empirical data are available and educational policies on this issue are inconsistent (Butler & Stevens, 2001). State educators and policymakers need to base their decisions about the type of accommodations they prefer to use for ELLs on empirical evidence. There are four major considerations for selecting appropriate accommodations: (a) effectiveness, (b) validity, (c) differential impact, and (d) feasibility. These four issues should be considered in combination and interactively. On the basis of current research, translating test items from English to other languages as an accommodation does not seem to be an effective strategy due to the fact that the subject is taught in English and reprocessing the learned information in Spanish is an unnecessary load for cognitive expression. Educators and policymakers should also consider the type of student group and differential impact when making accommodation choices. Commercial dictionaries are among the most commonly used accommodations; however, while they make test items with unfamiliar words more accessible, they can also provide content information, which leads to invalidity of the answer. Therefore, customized dictionaries have been found to help ELLs while not affecting the scores of non-ELLs. Empirical research studies have consistently showed positive results for modification of test language accommodation in narrowing the performance gap between ELLs and non-ELLs by removing low-frequency vocabulary and complex language structures that are irrelevant to the content knowledge being assessed. This accommodation strategy is effective and valid because it does not appear to affect non-ELL student performance.

Future research can build on research studies with a focus on linguistic
modification by reviewing and modifying test items through examining the effects of
different linguistic modification strategies that make unnecessarily complex and
construct-irrelevant test language accessible for ELLs. Another recommendation for
future studies is that ELLs should be taken into account at the beginning of the
development of a test and not as an afterthought. Often, test developers norm the test on
native English-speaking students and try to revise it to avoid language bias by making
adjustments later. Evidence from empirical research included suggestions that clear test
language without unnecessary complexity should not be an adaptation, but a valid and
reliable test practice.

Furthermore, empirical research on accommodations for kindergarten to Grade 12
ELLs is limited making anecdotal and best practice accommodation perceptions of
educational practitioners often the only resources available for them to make decisions
about accommodations. There is a dire need to develop and empirically test new and
innovative assessment strategies that can provide proven effectiveness and validity.
References


Collier, V. P. (1987). Age and rate of acquisition of second language for academic


Appendix A

Original and Linguistically Modified Test Item
1. What information led scientists to discover that the Earth’s ice cover is diminishing rapidly?
   A. the discovery of the greenhouse effect by a Swedish scientist about 30 years ago
   B. the fact that mountainous areas of the world are experiencing more precipitation
   C. the discovery by the British that the ice shelf in Antarctic is stable and predictable
   D. a discovery of open water at the North Pole and evidence from two scientific studies

2. What effects would the rising sea level have on Bangladesh?
   A. The rainfall would increase and the ice mass would diminish.
   B. Mountainous areas would see less rainfall, depleting summer water reserves.
   C. The country would lose much of its farmland forcing residents to move inland.
   D. Temperatures of low-lying areas would rise, forcing people to move to cooler coastal areas.

3. What evidence best supports the author’s claim that people have been aware of the changes in the earth’s climate for a very long time?
   A. Over the last century, sea levels rose significantly in all parts of the world.
   B. The glaciers in the Alps have shrunk about 40 percent in the last 50 years.
   C. The warmest 23 years have all taken place in approximately the last 30 years.
   D. In the early 20th century, a scientist warned of the risks of increased CO₂ levels.

4. What is most likely the author’s purpose in writing this article?
   A. to explain how scientists became aware of the reduction of the polar ice cover
   B. to convince readers to help fund research to learn the causes of the melting ice cap
   C. to illustrate effects of global warming on the world, especially the Asian population
   D. to convince readers that the melting ice cap illustrates a need for alternate sources of power

5. What evidence supports the idea that the glaciers in Europe’s Alps may disappear over the next fifty 50 years?
   A. The shrinkage of ice masses always follows a consistent pattern.
   B. In the past 150 years, glaciers have shrunk to almost one half their
original size.
C. The shrinking ice mass in the Himalayas has advanced by 36 meters since 1992.
D. The U.S. Geological Survey projects that the remaining glaciers will disappear soon.

6. How will the melting snow/ice masses in the Himalayas affect the water supply of Asia?
   A. The floodplains will overflow and people will need to learn flood control farming.
   B. As the snowmelt that feeds rivers diminishes, many countries will become even drier.
   C. The amount of precipitation will increase, causing more flooding during the dry season.
   D. As the snowmelt decreases, area rivers will change course, no longer flowing into countries that need water the most.

7. According to the article, how is the Antarctic different from the North Pole?
   F. It is covered by solid ground.
   G. It is surrounded by ice shelves.
   H. It is studied by climate scientists.
   I. It is unaffected by rising CO₂ levels.

8. The author organizes the article by
   A. explaining the effects of global warming and then listing its causes.
   B. presenting a theory about melting ice and then supporting it with examples.
   C. describing an arctic hike and then showing how it relates to global warming.
   D. describing evidence of melting ice and then giving explanations and solutions.

9. Read this sentence from the article.
The thinning and shrinkage has reduced the Arctic Ocean ice mass by nearly half.
Which of the following sentences uses the word mass with the same meaning as in the sentence above?
   A. Mass of excited fans crowded around the movie star’s limousine.
   B. Weighted down by its enormous mass, the ocean liner sank rapidly.
   C. Hearing news of the sale, shoppers began to mass outside the store at 5 A.M.
   D. Scientists claimed the mass of the unexplored planet and divided it into colonies.
1. What information helped scientists to discover that the Earth’s ice cover is disappearing quickly?
   E. the discovery of the greenhouse effect by a Swedish scientist about 30 years ago
   F. mountainous areas of the world are experiencing more precipitation
   G. the discovery by the British that the ice shelf in Antarctic is stable and predictable
   H. a discovery of open water at the North Pole and evidence from two scientific studies

2. What would happen to Bangladesh if the sea continued to rise?
   E. The rainfall would increase and the ice mass would disappear.
   F. There would be less rainfall in mountainous areas that will empty summer water reserves.
   G. The country would lose much of its farmland and force residents to move inland.
   H. Temperatures of low-lying areas would rise, and force people to move to cooler coastal areas.

3. What is the evidence that show that people have known the changes in the earth’s climate for a very long time?
   E. Over the last century, sea levels rose significantly in all parts of the world.
   F. The glaciers in the Alps have gotten smaller about 40 percent in the last 50 years.
   G. The warmest 23 years happened in approximately the last 30 years.
   H. In the early 20th century, a scientist warned of the risks of increased CO₂ levels.

4. What is the author’s purpose in writing this article?
   E. to explain how scientists became aware that the polar ice cover is getting smaller.
   F. to convince readers to help donate money to research to learn the reasons for the melting ice cap
   G. to show effects of global warming on the world, especially the Asian population
   H. to convince readers that the melting ice cap shows a need for other sources of power

5. Which one of these is the evidence that the glaciers in Europe’s Alps may disappear over the next 50 years?
   E. Ice masses always has a consistent pattern.
   F. In the past 150 years, glaciers became smaller to almost one half their
The ice mass in the Himalayas got smaller by 36 meters since 1992. The U.S. Geological Survey predicts that the remaining glaciers will disappear soon.

6. What will happen to the water supply of Asia if snow and ice masses continue to melt?
   E. The floodplains will overflow and people will need to learn flood control farming.
   F. As the snowmelt that feeds rivers disappears, many countries will become even drier.
   G. The amount of precipitation will increase and cause more flooding during the dry season.
   H. As the snowmelt decreases, area rivers will change direction, no longer flowing into countries that need water the most.

7. According to the article, what is the difference between the Antarctic and the North Pole?
   J. It has solid ground.
   K. Ice shelves surround it.
   L. Climate scientists study it.
   M. It is not affected by rising CO\textsubscript{2} levels.

8. The author organizes the article by
   E. explaining the effects of global warming and then listing its reasons.
   F. showing a theory about melting ice and then supporting it with examples.
   G. describing a walk in the Arctic and then showing how it relates to global warming.
   H. describing evidence of melting ice and then giving explanations and solutions.

9. Read this sentence from the article.
   **The thinning and shrinkage has reduced the Arctic Ocean ice mass by nearly half.**
   Which of the following sentences uses the word mass with the same meaning as in the sentence above?
   E. A mass of excited fans crowded around the movie star’s limousine.
   F. Heavy with its enormous mass, the ocean liner sank rapidly.
   G. Hearing there is a sale, shoppers began to mass outside the store at 5 A.M.
   Scientists claimed the mass of the unexplored planet and divided it into colonies

Appendix B

Interview Questions
**Interview Questions**

**Experimental group**
1. You speak Spanish at home and the language used at school is English. How do you feel about this?
2. When you are listening to the lesson your teacher is teaching, what helps you most in understanding the lesson?
3. Do you feel that you know the content of a lesson but cannot answer the questions when you take a test on it? Why?
4. Can you usually show what you know and can do in a test in English?
5. Let’s say you studied very hard for a test. What do you worry about the most before you take a test in English?
6. What do you do when you don’t understand the question in a test?
7. Can you find the meaning of unfamiliar English words in the test when you look for them in the dictionary?
8. You took two tests on the …………..unit you studied. Can you tell me if you noticed a difference between the two tests? What do you think was the difference?
9. Why do you think x was easier/more difficult than y?
10. Which one these is most helpful for you when you take a test? Why?
   a. Extended time/dictionary use
   b. Easy to understand questions and multiple-choice answers

**Control group**
1. You speak Spanish at home and the language used at school is English. How do you feel about this?
2. When you are listening to the lesson your teacher is teaching, what helps you most in understanding the lesson?
3. Can you usually show what you know and can do in a test in English?
4. Let’s say you studied very hard for a test. What do you worry about the most before you take a test in English?
5. What do you do when you don’t understand the question in a test?
6. Can you find the meaning of unfamiliar words in the test when you look for them in the dictionary?
7. You took two tests on the ….unit. You were allowed to use dictionary and extended time. If you were allowed only 1 accommodation (extended time or dictionary use) which one of these would be the most helpful for you? Why?
8. (If the answer is dictionary); Can you always find meanings of unfamiliar words in the dictionary?
9. Do you feel that you know the content of a lesson but cannot answer the questions when you take a test on it? Why?
10. What would be most helpful for you to show what you know in tests administered in English?
Appendix C

Linguistic Modification Guide
Developing a Linguistically Modified Assessment Guidelines adapted from Linguistic Modification In LEP Partnership (Abedi & Sato, 2007 pp. 81-84)

<table>
<thead>
<tr>
<th>STEP I: Definition of the domain and constructs</th>
<th>STEP II: Definition of the population and their access needs</th>
<th>Step III: Linguistic Modification Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Purpose of the assessment</td>
<td>▪ State the characteristics and access needs of ELLs</td>
<td>▪ Categorize content/items based on their eligibility to be linguistically modified or not. (* Eligibility is determined based on the construct-relevancy of the language)</td>
</tr>
<tr>
<td><em>The assessment results will be used for the following reasons...</em></td>
<td><em>The ELLs tested have these characteristics:........</em></td>
<td></td>
</tr>
<tr>
<td>▪ State the domain to be assessed.</td>
<td><em>Their access needs are:............................................</em></td>
<td></td>
</tr>
<tr>
<td><em>The results of the assessment will be used for the following reasons:....</em></td>
<td>Context: Words, Phrases, Sentences: Format/ Style:</td>
<td></td>
</tr>
<tr>
<td>▪ Evaluate. Does the assessed domain match the stated purpose of the assessment?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ State the specific content and skills to be measured.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>The assessment will measure the following specific constructs related to the domain:......</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ State content related language.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>The following vocabulary and terminology are specific to the grade-level content assessed:.......</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Apply linguistic modification guidelines for context, graphics, vocabulary/wording, sentence structure, format/style</td>
<td></td>
<td></td>
</tr>
<tr>
<td>▪ Evaluate the linguistically modified items. Check to see if the linguistically modified version of the test item provides linguistic access to ELLs in terms of context, graphics, vocabulary/wording, sentence structure, format/style</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Significant Statements in Theme 7: Use of Dictionaries in Tests
## Significant Statements in Theme 7: Use of Dictionaries in Tests

<table>
<thead>
<tr>
<th>Student</th>
<th>Significant statement</th>
<th>Subtheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Yes I find them, but sometimes later I find out it wasn’t the right meaning. If I find the meaning, I sometimes don’t understand the long sentence it is used in.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>A</td>
<td>I look in the dictionary, but in FCAT, they don’t let us use our own dictionaries and I am not used to the ones they give us so it becomes more difficult to find the meanings.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>A</td>
<td>I worry that I will pick the wrong answer because I can’t find the meaning of words in my dictionary.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>A</td>
<td>I look for it in the dictionary and try to find the right meaning, but sometimes the word is too difficult and I don’t even understand it in Spanish.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>B</td>
<td>I look in the dictionary, but sometimes it’s very confusing because there are more than one meaning and I don’t know which one is correct. So I just pick one without knowing for sure.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>B</td>
<td>But if I have time after I look at the words in my dictionary, I wish difficult words in the tests had meanings in Spanish right next to them so I can understand and move on. Sometimes the meaning of an English word is a word I don’t know in Spanish either. If they used an easier word, I would understand it I think.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>A</td>
<td>I look for it in the dictionary and try to find the right meaning but sometimes the word is too difficult and I don’t even understand it in Spanish.</td>
<td>Difficult to find meaning</td>
</tr>
<tr>
<td>C</td>
<td>They allow us to use dictionaries, but that doesn’t always help.</td>
<td>Does not always help</td>
</tr>
<tr>
<td>D</td>
<td>I find it sometimes but some words and phrases are not in the dictionary so you have to make a guess.</td>
<td>Some words and phrases not found</td>
</tr>
<tr>
<td>A</td>
<td>I know I would lose so much time looking in the dictionary so many times. Maybe I wouldn’t even need to use it.</td>
<td>Time</td>
</tr>
<tr>
<td>C</td>
<td>It’s hard to use the dictionary; I just started learning it. I wish we could practice with the teacher and learn how to use it fast because we worry about losing time and not finishing on time.</td>
<td>Time</td>
</tr>
<tr>
<td>C</td>
<td>I will not have enough time if I search the words in the dictionary.</td>
<td>Time</td>
</tr>
<tr>
<td>D</td>
<td>Dictionary is useful. I use dictionary a lot. If you don’t know how to use a dictionary fast, then you lose too much time and even extra time isn’t enough.</td>
<td>Time</td>
</tr>
</tbody>
</table>