
Theses and Dissertations

2013

Testing accommodations for ELL students on an achievement test battery

Lori Lee Dockery
University of Iowa

Copyright 2013 Lori Dockery

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/2477>

Recommended Citation

Dockery, Lori Lee. "Testing accommodations for ELL students on an achievement test battery." PhD (Doctor of Philosophy) thesis, University of Iowa, 2013.
<http://ir.uiowa.edu/etd/2477>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Educational Psychology Commons](#)

TESTING ACCOMMODATIONS FOR ELL STUDENTS
ON AN ACHIEVEMENT TEST BATTERY

by
Lori Lee Dockery

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychological and Quantitative Foundations
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Timothy Ansley

How well students perform on standardized tests can affect their educational paths and the rest of their lives. In addition, students' performances on state assessments will affect their schools due to the No Child Left Behind Act. For English language learners (ELLs), the success on tests may be diminished due to their inability to completely understand what they are reading on a test. Because ELLs are a growing proportion of the population and have greater risk of not performing well in school, dropping out, and not moving on to have a job, it is very important to address their educational performance. To alleviate the difficulty of not being able to understand English competently, various testing accommodations can be given. The purpose of this study was to describe how different types of accommodations are being distributed, particularly among ELL students, on an achievement test battery. Several variables were examined to assess whether they related to which accommodation would be assigned to an individual student.

This study used data from a recent 2010 national standardization of an achievement test battery, which sampled 33,226 students from grades 3-8 across the U. S. In addition to the tests, students completed a survey which asked for such information as gender, home language, and ethnicity, and test administrators reported on that same survey which testing accommodations students were given, whether students were migrants, whether students participated in a free or reduced-price lunch program, and whether students were ELL students. These variables, along with students' Reading, Math, and Science scores were used to describe the groups of students given each accommodation.

Five testing accommodations were reported by the students – giving students extended time to take the test, allowing students to use a word-to-word dictionary, reading parts of the test aloud to the students, repeating instructions, and having the test administered by an ELL teacher. Of all these, the group that was most similar to the entire sample on the test results for the Reading, Math, and Science was the group given

dictionaries, a group predominantly comprised of Asian students. The other testing accommodation groups had much lower percentile ranks on average. ELL students were largely Hispanic, spoke Spanish at home, and had lower percentile ranks on the Reading, Math, and Science tests than non-ELL students. Although the majority of ELL students in this sample was not free or reduced-price lunch eligible, there was a higher percentage than was found in the sample as a whole.

It was found that several variables were potentially important in how testing accommodations were distributed among students, including grade level, ethnicity, home language, and socioeconomic status (SES). Variables which did not seem important were gender and school. Home language and SES were also important in testing performance, with low SES students performing much worse on average than the sample as a whole. Students who had Spanish as their home language did not perform as well on the tests as students whose home language was English or another language. Lastly, a gap analysis using effect sizes showed some evidence for the gap between ELL and non-ELL students being larger in higher grade levels.

Abstract Approved: _____
Thesis Supervisor

Title and Department

Date

TESTING ACCOMMODATIONS FOR ELL STUDENTS
ON AN ACHIEVEMENT TEST BATTERY

by
Lori Lee Dockery

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychological and Quantitative Foundations
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Associate Professor Timothy Ansley

Copyright by
LORI LEE DOCKERY
2013
All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Lori Lee Dockery

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Psychological and Quantitative Foundations at the May 2013 graduation.

Thesis Committee: _____
Timothy Ansley, Thesis Supervisor

Stephen Dunbar

Lia Plakans

Walter Vispoel

Catherine Welch

ACKNOWLEDGMENTS

I am very grateful to my advisor, Dr. Tim Ansley, for providing feedback on the many versions of this paper. In addition, I want to thank the ITP staff who provided me with support with my dataset, Tim Burden and Josh Tudor.

ABSTRACT

How well students perform on standardized tests can affect their educational paths and the rest of their lives. In addition, students' performances on state assessments will affect their schools due to the No Child Left Behind Act. For English language learners (ELLs), the success on tests may be diminished due to their inability to completely understand what they are reading on a test. Because ELLs are a growing proportion of the population and have greater risk of not performing well in school, dropping out, and not moving on to have a job, it is very important to address their educational performance. To alleviate the difficulty of not being able to understand English competently, various testing accommodations can be given. The purpose of this study was to describe how different types of accommodations are being distributed, particularly among ELL students, on an achievement test battery. Several variables were examined to assess whether they related to which accommodation would be assigned to an individual student.

This study used data from a recent 2010 national standardization of an achievement test battery, which sampled 33,226 students from grades 3-8 across the U. S. In addition to the tests, students completed a survey which asked for such information as gender, home language, and ethnicity, and test administrators reported on that same survey which testing accommodations students were given, whether students were migrants, whether students participated in a free or reduced-price lunch program, and whether students were ELL students. These variables, along with students' Reading, Math, and Science scores were used to describe the groups of students given each accommodation.

Five testing accommodations were reported by the students – giving students extended time to take the test, allowing students to use a word-to-word dictionary, reading parts of the test aloud to the students, repeating instructions, and having the test

administered by an ELL teacher. Of all these, the group that was most similar to the entire sample on the test results for the Reading, Math, and Science was the group given dictionaries, a group predominantly comprised of Asian students. The other testing accommodation groups had much lower percentile ranks on average. ELL students were largely Hispanic, spoke Spanish at home, and had lower percentile ranks on the Reading, Math, and Science tests than non-ELL students. Although the majority of ELL students in this sample was not free or reduced-price lunch eligible, there was a higher percentage than was found in the sample as a whole.

It was found that several variables were potentially important in how testing accommodations were distributed among students, including grade level, ethnicity, home language, and socioeconomic status (SES). Variables which did not seem important were gender and school. Home language and SES were also important in testing performance, with low SES students performing much worse on average than the sample as a whole. Students who had Spanish as their home language did not perform as well on the tests as students whose home language was English or another language. Lastly, a gap analysis using effect sizes showed some evidence for the gap between ELL and non-ELL students being larger in higher grade levels.

TABLE OF CONTENTS

LIST OF TABLES	vii
CHAPTER ONE: INTRODUCTION.....	1
ELL Statuses	2
Description of ELL Students	2
ELL Student Performance	3
Accommodations	5
Accommodation Selection.....	6
Previous research	7
Purpose of the study.....	8
CHAPTER TWO: LITERATURE REVIEW	11
Accommodation Research	11
Accommodation vs Modification	13
Taxonomy	14
Presentation Accommodations	15
Dictionaries.....	15
Computerized Testing	18
Linguistic Modification	19
Translated Test	22
Read Aloud.....	23
Repeating Instructions	28
Extended Oral Presentation	24
Response Accommodations.....	25
Setting Accommodations.....	26
Timing/Scheduling Accommodations	28
Extended Time.....	26
Important Considerations for Deciding on Accommodations	29
Ethnicities in Previous Research	30
Grade Levels in Previous Research	30
Assessments Used in Previous Research.....	31
Locations of Previous Research.....	34
Ethnicities	34
Poverty.....	36
Percentage of ELL Students	37
Operational Testing	38
Summary	40
CHAPTER THREE: METHODS	46
Sample and Procedure	46
Participants	47
Measures	48
Rationale for Chosen Study Variables.....	50
Analysis	52

CHAPTER FOUR: RESULTS	55
Environment	55
Individual Characteristics	62
Test Scores	77
High vs Low Stakes	90
Gap Comparison	94
Summary of Accommodation Groups	94
 CHAPTER FIVE: DISCUSSION.....	 101
Summary	104
Limitations	104
Future Research	106
Final Comments.....	106
 REFERENCES	 108
 APPENDIX	
STUDENT/TEST ADMINISTRATOR SURVEY	119

LIST OF TABLES

Table 1. Percentage of Students in Ethnic Groups by State	35
Table 2. Percentage of Students Eligible for Free or Reduced-Price Lunch by State.....	37
Table 3. Percentage of Students Receiving ELL Services by State.....	38
Table 4. Summary of Previous Research	42
Table 5. Number of Test Items per Grade for Each Content Area	50
Table 6. Number of ELL Students Receiving Accommodations per School	56
Table 7. Total ELL Students Receiving Accommodations by Number of ELL Students in Schools.....	58
Table 8. Total ELL Students Receiving Accommodations by Proportion of ELL Students in Schools.....	59
Table 9. Grade Levels for Sample, ELL Students, and Accommodations Groups.....	59
Table 10. Gender for Sample, ELL Students, and Accommodations Groups	61
Table 11. Ethnicities for Sample, ELL Students, and Accommodations Groups.....	64
Table 12. Ethnic Group Comparison between Study Sample and the United States.....	65
Table 13. Ethnicities for Sample, ELL Students, and Accommodations Groups with No Overlap.....	70
Table 14. Home Languages for Sample, ELL Students, and Accommodations Groups.....	73
Table 15. Free or Reduced-Price Lunch Eligibility for Sample, ELL Students, and Accommodations Groups.....	76
Table 16. Reading National Percentile Ranks for Sample, ELL Students, and Accommodations Groups.....	78
Table 17. Math National Percentile Ranks for Sample, ELL Students, and Accommodations Groups.....	80
Table 18. Science National Percentile Ranks for Sample, ELL Students, and Accommodations Groups.....	81
Table 19. Reading, Math, and Science National Percentile Ranks for ELL Students and ELL Students with Testing Accommodations	82
Table 20. Reading National Percentile Ranks by Home Language	83

Table 21. Math Percentile Ranks by Home Language.....	84
Table 22. Science National Percentile Ranks by Home Language for ELL Students and ELL Students with Testing Accommodations	85
Table 23. Reading National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations	86
Table 24. Math National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations	86
Table 25. Science National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations	87
Table 26. Reading, Math, and Science National Percentile Ranks for FRPL Eligible and Non-Eligible Students for ELL Students and ELL Students with Testing Accommodations	89
Table 27. Minimum and Maximum Percentages of Proficiency and Above on State Assessments for Overall Samples.....	91
Table 28. Minimum and Maximum Percentages of Proficiency and Above on State Assessments for ELL Samples	92
Table 29. Minimum and Maximum Effect Sizes and Average Effect Sizes for the Gap Between ELL and Overall Percentages of Proficiency and Above on State Assessments	96
Table 30. Minimum and Maximum Mean Scores on NAEP for Overall Samples.....	97
Table 31. Minimum and Maximum Mean Scores on NAEP for ELL Samples	97
Table 32. Minimum and Maximum Effect Sizes and Average Effect Sizes for the Gap Between ELL and Overall Mean Scores on the NAEP.....	98
Table 33. Effect Sizes and Confidence Intervals for the Gap between ELL and Overall Average Percentile Ranks of the Study Sample	99

CHAPTER ONE – INTRODUCTION

The No Child Left Behind Act of 2001 (NCLB; 2002) was designed to encourage schools to improve the achievement of all students. This legislation provided mechanisms for motivating school personnel to ensure that all students are reaching proficiency by the year 2014. These mechanisms included requiring states to set percentages of students expected to be proficient, as defined by the state, that steadily increase to 100%, and allowing students to attend different schools or reducing funds if a school was found to be doing poorly according to NCLB requirements. NCLB required that each subgroup with more than some specified minimum number of students, such as each group of students of a certain race, also perform at the same level of proficiency required for the overall school.

Among the more clearly defined goals provided by NCLB were expectations for English Language Learners (ELLs) (NCLB, 2002). Students are identified as ELL by taking a state-selected English proficiency test and scoring low enough to be determined to be limited in listening, speaking, reading, and/or writing English. Although they do not have proficiency in English, ELLs are expected to progress towards full proficiency like any other subgroup, making the scores of ELL students impact the overall status of a school's achievement levels in math, science, and reading.

In addition, ELL students were to be analyzed as a subgroup. An interesting phenomenon related to the progress of this subgroup is that once ELL students reach proficiency in English, their status will change, no longer counting toward the ELL subgroup scores (Abedi, 2004). Because ELL students typically score lower than non-ELL students in all areas (Abedi, 2002) and increase proficiency in other academic areas only as they also gain proficiency in English (Abedi & Lord, 2001), a subgroup which only consists of students not proficient in English will likely continue to have low scores.

The federal government provided some relief for this issue in 2006 (Title I, 2006) by allowing recent ELL arrivals to the United States to skip the reading assessment their first year. Additionally, ELL students can be allowed to take the reading assessment in their first language for the first three years they are in the United States. They are still required to take the math and science assessments in English, however. NCLB makes it very important to schools to provide services which will improve ELL students' test scores. Even without the legislation, though, it is important for all students to succeed to their fullest potential.

ELL Statuses

ELL is often used synonymously with Limited English Proficient (LEP) (Shepard, Taylor, & Betebenner, 1998), including in the NCLB legislation (2002). Although both ELL and LEP may refer to the same group of students, many in the field of education feel that describing students as "limited" gives a negative undertone, so they prefer using ELL as a description, with "learning" having more positive connotations (Iowa Department of Education, 2007).

ELL students are often categorized as fluent English proficient (FEP) once they have passed an English proficiency test or somehow similarly proved their level of proficiency with English (Butler & Castellon-Wellington, 2000). Although a student may have been judged to be proficient in English, and be reclassified as FEP, learning of English can continue and his/her ability level in English will keep increasing for the next few years, albeit at a reduced rate (Hakuta, Goto, & Witt, 2000).

Description of ELL Students

The majority of ELL students in the United States are born in the United States (Batalova, Fix, Murray, 2007; NCES, 2004). There are approximately equal numbers of boys and girls across ELL statuses (Abedi, 2008), although some research suggests that girls are more likely to be reclassified as English proficient than boys (Grissom, 2004).

ELL students as a whole are more likely to be from lower-income families (Abedi, 2002; Abedi, Leon, & Mirocha, 2000; Abedi and Lord, 2001; Batalova, Fix, & Murray, 2007; Collier, 1987) and have parents with lower education levels (Abedi, 2002). In one study, 41% of ELL students were of high socioeconomic status (SES), while 72% of non-ELL students were of high SES (Abedi and Lord, 2001). Grissom (2004) found that ELL students who are from high SES families also have the advantage of being more likely to achieve reclassification. However, Abedi (2008) found that the median time classified as ELL before being reclassified as English proficient was nearly the same when free and reduced-price lunch eligible ELL students were compared to non-eligible students.

The population growth of ELL students in the United States is far outpacing the growth of the overall population of students (Batalova, Fix, & Murray, 2007). The percentage of people between the ages of 5 and 24 who speak a language other than English at home rose from 9% to 17% between 1979 and 1999 (NCES, 2004). The percentage of White, non-Hispanic students is decreasing. The percentage was 57.1 during the 2005-2006 school year (Sable & Garofano, 2007), and 54.3 during the 2008-2009 school year (NCES, 2009). The majority of non-White students are Hispanic (21.2%), while the Black, non-Hispanic group is the other large non-White group (16.8%) (NCES, 2009).

ELL Student Performance

With linguistic, cultural, and socioeconomic differences between ELL students and non-ELL students, there are bound to be educational impacts. ELL students consistently score lower than non-ELL students on tests in math, science, and reading (Abedi, 2002; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Abedi, Leon, & Mirocha, 2000; Abedi, Leon, & Mirocha, 2003; Abedi & Lord, 2001; Abedi, Lord, & Hofstetter, 1998; Abedi, Lord, Kim, & Miyoshi, 2001; Abedi, Lord, & Plummer, 1997;

Butler & Castellon-Wellington, 2000; Hafner, 2001; Hofstetter, 2003; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Rivera & Stansfield, 2003; Shepard, Taylor, & Betebenner, 1998; Young, Cho, Ling, Cline, Steinberg, & Stone, 2008). The performance gap is larger in areas which require better aptitude with the English language, such as reading (Abedi, 2002; Abedi, Leon, & Mirocha, 2000; Abedi, Leon, & Mirocha, 2003). The gap is also larger as grade level increases (Abedi, Leon, & Mirocha, 2000; Butler & Castellon-Wellington, 2000; Hakuta, Goto, & Witt, 2000). This may be because what is being taught becomes more complex, more reliant on language to be learned. Or it may be that ELL students, although they are grasping English more each year, have lost out on the foundation that is the basis for additional learning.

There is variation among ELL students that can contribute to differences in how they perform on academic tests. One variable to consider is how much English is spoken in the home. Those who only speak a language other than English in the home do not do as well on academic tests as those who speak English as well as another language at home (Abedi, Lord et al, 2001; Abedi, Lord, & Plummer, 1997). Those who speak a language other than English in the home, whether they do so exclusively or also speak English, also do not do as well as those who only speak English in the home on other educational variables, such as staying in school, as well as economic variables, such as being able to obtain a job (NCES, 2004). Even among students who are designated as ELL, those who are more proficient in English do better than those with lower proficiency (Abedi & Lord, 2001).

The majority of ELL students speak Spanish in the home (Batalova, Fix, & Murray, 2007; NCES, 2004). Those ELL students whose other language is Spanish have been found to particularly lag behind non-ELL students in test performance (Myers & Milne, 1988). They are also more likely to drop out of school and repeat grades (NCES, 2004). ELL students who speak Spanish at home are less likely than other ELL students to achieve reclassification (Abedi, 2008; Grissom, 2004). Students who do not become

proficient in English while in school, especially Spanish speakers, are much less likely than non-ELL students to complete high school, go on to college, be from middle or high income households, and have jobs (NCES, 2004).

In contrast, it has been suggested by earlier research that those students who speak another language besides English or Spanish at home perform better than students who only speak English at home (Myers & Milne, 1988). For example, those who also speak Chinese, Italian, French, or German at home are likely to have higher math scores. Because ELL students are becoming a larger proportion of the population, and the majority are Hispanic students who are unlikely to succeed beyond school if they do not do well in school, it is important to focus more efforts on providing ELL students with tools that will assist them in every way to succeed in school.

Accommodations

The goal of an accommodation is to alleviate barriers that do not pertain to what is being tested, thereby increasing the validity of the scores (Rivera & Stansfield, 1998). For ELL students, that means being able to assess their skills in math or science, for example, without their language difficulties masking their true ability levels. An appropriate accommodation should increase an ELL student's score, but, if implemented, would not change a non-ELL student's score (Abedi, Hofstetter, & Lord, 2004).

There are two main categories of accommodations – changes to the test and changes to the testing procedure (Butler & Stevens, 1997). Accommodations which are categorized as changes to the test include translating it into the student's native language, using easier vocabulary, making the language less complex, adding visual aids, adding a glossary in the native language or English, and adding more sample items at the beginning of the test (Butler & Stevens, 1997). Accommodations which are categorized as changes to the testing procedure include giving the student extended time to complete the test, adding more breaks during the test administration, breaking the testing into

sessions that can be administered across a greater time span, giving directions in the student's native language, administering the test in smaller groups, separating the ELL students to their own testing room, allowing the use of dictionaries, reading the instructions or questions aloud in English, and having the administrator provide further explanations for the test instructions (Butler & Stevens, 1997).

Accommodation Selection

In order to gain a better understanding of how testing accommodations are selected by states for the ELLs, state department of education websites were searched. Relevant documentation was found for Washington D.C. and 46 of the states. While some documents indicated that guidance was minimal or was left to the local level, most states were very thorough in describing who was to be on the team making the decision and what factors were to be considered. Overall, the consensus was that team members involved in the accommodation decision making should be knowledgeable about state content standards and assessments, as well as acquainted with the student. The team could include ELL or bilingual teachers, general education teachers, the school/district test coordinator, the ELL/bilingual coordinator, counselors, and school administrators. Students and parents could also be team members if it seemed appropriate.

There were many factors that the documents said were to be taken into account. The most popular one was the student's level of proficiency of English. After that, the perceptions of how well the accommodation worked by the student, parents, teachers, and specialists; the student's academic strengths and weaknesses; the results when a comparison was made between the student using the accommodation and not using it; the opportunity for the student to learn how to use the accommodation; the student's level of literacy in the native language; the student's difficulty in using the accommodation; the student's willingness to learn to use the accommodation; the student's grade/age/maturity; and the length of time the student had been in the U.S. were all fairly

popular. There were emphases in nearly all documents that the accommodation should be fit to each individual student on a case-by-case basis and that the student should have the opportunity to practice using the accommodation in the classroom before using it on the state assessment. Many also mentioned that the decision-making process was recursive and that the decisions should be re-evaluated at least annually.

Previous Research

Much of the previous research on how testing accommodations affect ELL students' test scores was done in California using items from the National Assessment of Educational Progress (NAEP) in an experimental setting (Abedi, Courtney, & Leon, 2003; Abedi, Hofstetter, Baker, & Lord, 2001; Abedi & Lord, 2001; Abedi, Lord et al, 2001; Abedi, Lord, & Plummer, 1997; Abedi et al, 2005; Hofstetter, 2003). There has been some research in other U.S. locations (Kopriva et al, 2007; Albus, Thurlow, Liu, & Bielinski, 2005; Rivera & Stansfield, 2003; Liu, Anderson, Swierzbin, & Thurlow, 1999; Shepard, Taylor, & Betebenner, 2008), using other tests (Albus et al, 2005; Castellon-Wellington, 2000; Hafner, 2001; Liu et al, 1999; Rivera & Stansfield, 2003; Shepard, Taylor, & Betebenner, 1998), and in operational situations (Rivera & Stansfield, 2003; Shepard, Taylor, & Betebenner, 1998). Overall, the research has not consistently shown that accommodations raise ELL students' scores while not raising non-ELL students' scores. It is possible that studies that showed significant results differed on some variable, such as SES or race, when compared to a study using the same accommodation that did not show any significant results. The current research adds further perspective to the existing research through the use of a large, unique data set.

The majority of the studies gathered data using procedures different from operational testing administrations (Abedi, Courtney, & Leon, 2003; Abedi et al, 2005; Abedi, Hofstetter et al, 2001; Abedi & Lord, 2001; Abedi, Lord et al, 2001; Abedi, Lord, & Plummer, 1997; Albus et al, 2005; Bensoussan, Sim, Weiss, 1984; Castellon-

Wellington, 2000; Hafner, 2001; Hofstetter, 2003; Liu et al, 1999; Nesi & Meara, 1991). They used such methods as randomly assigning students to accommodations, changing the number of items or duration of tests, using older versions of tests, piecing together tests from multiple sources, creating new test questions, and including interviews. With normal testing conditions, the students' motivation to do well is likely going to be higher, and the students are more likely to have accommodations which have been individually assigned in a tailored fashion.

Purpose of the Study

Given the current NCLB regulations, it is very important to schools that all students score at their full potential so that the school does not suffer negative consequences. Thus, it behooves them to find appropriate ways to help ELL students overcome the language barrier. Although there is some previous research in the area of accommodations, this study has some unique aspects which add to the research available and will hopefully help future research to be more useful for those making accommodations decisions for ELL students.

There are three main differences between previous research and the current study: the way accommodations are assigned to students, the assessment used, and how high the stakes are. In most of the previous research on testing accommodations, students were assigned to accommodations randomly. In most states, however, accommodations are supposed to be assigned on a case-by-case basis to make certain that the accommodation will be helpful for the students. In order to find out whether accommodations are indeed being individually assigned and that it is not the case that students are being given a blanket accommodation, this study describes the students being given accommodations using several variables that have been found to be relevant.

This study differs from most previous studies by using a different assessment. While the NAEP, the most commonly used assessment in the previous testing

accommodation research, is a well-known national measure, it is not currently used by any state as a formal evaluation or diagnostic tool. There have been some studies which did not use the NAEP, but these also did not involve widely-used tests. A standardized achievement test battery, used by many states for providing information to evaluate student achievement and diagnose areas of academic strength and weakness, was used in this study. The standardized achievement battery items also differ from those used in some of the studies in that they are entirely multiple choice questions.

The third main difference between this study and most previous research is that the assessment was used within the context of a norming study. The stakes on a norming study are not as high for schools or students as on a state assessment used for NCLB accountability, but the assessment was still given in a standardized situation with questions and timing exactly like the operational setting. This context could appear quite similar to the high stakes environment of the state assessment, potentially eliciting a similar level of motivation and performance. In most of the previous research, however, the number of items and amount of time were kept to a minimum and it is not clear how well the situations mimicked an operational setting. It is most likely that the research settings of those studies made it apparent that there were no stakes, so the students potentially would not try as hard to answer questions correctly, which could have affected the results.

This study is guided by the following research questions.

- 1) What is the relationship between socioeconomic status, migrant status, ethnicity, gender, grade level, home language, and which accommodations the ELL students receive?
- 2) To what extent do Math, Science, and Reading scores vary among ELL students receiving different types of accommodations?

- 3) How does the gap between ELL and non-ELL performance on a norming study compare to low stakes (NAEP) and high stakes (NCLB state assessments) ELL achievement gaps?

CHAPTER TWO – LITERATURE REVIEW

In chapter one of this paper, there was a discussion of the reasoning behind giving English language learners (ELLs) accommodations on tests and how this study adds to the body of research in this area. In this chapter, there is an exploration of the existing research on testing accommodations for ELL students, providing a background for the current research. Different accommodations are covered, including allowing dictionaries, computerized testing, extending the administration time, linguistic modification, and translating the test. Also covered is how different ethnicities, grade levels, assessments, and locations have been examined in previous research.

Accommodation Research

Most researchers have found no significant differences in student performance when there was an accommodation and when there was no accommodation (Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997; Bensoussan, Sim, & Weiss, 1984; Castellon-Wellington, 2000; Kopriva et al, 2007; Nesi & Meara, 1991; Rivera & Stansfield, 2003; Young et al, 2008) or that the accommodation provided similar benefits to both ELL and non-ELL students (Hafner, 2001; Shepard, Taylor, & Betebenner, 1998).

Of the studies examined in this literature review, there were a number of accommodations used:

- allowing the use of an English dictionary (Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Courtney, & Leon, 2003; Albus et al, 2005; Bensoussan, Sim, Weiss, 1984; Nesi & Meara, 1991);
- using a picture dictionary (Kopriva et al, 2007);
- using a computerized version of the test (Abedi, Courtney, & Leon, 2003);

- allowing extra time (Abedi, Hofstetter et al, 2001; Abedi, Courtney, & Leon, 2003; Castellon-Wellington, 2000; Hafner, 2001; Shepard, Taylor, & Betebenner, 1998);
- small group administration (Abedi, Courtney, & Leon, 2003; Shepard, Taylor, & Betebenner, 1998);
- linguistic modification (Abedi, Hofstetter et al, 2001; Abedi et al, 2005; Abedi & Lord, 2001; Abedi, Lord, and Hofstetter, 1998; Abedi, Lord, & Plummer, 1997; Hofstetter, 2003; Rivera & Stansfield, 2003);
- adding a glossary to the test (Abedi, Hofstetter et al, 2001; Abedi, Courtney, & Leon, 2003);
- allowing the use of a bilingual dictionary (Abedi et al, 2005; Bensoussan, Sim, Weiss, 1984);
- adding a bilingual glossary to the test (Kopriva et al, 2007);
- putting translations of key words in the margins (Abedi, Lord et al, 2001);
- using a bilingual test (Liu et al, 1999);
- using a test translated into the student's first language (Hofstetter, 2003; Shepard, Taylor, & Betebenner, 1998);
- having the test questions read aloud (Castellon-Wellington, 2000; Kopriva et al, 2007; Shepard, Taylor, & Betebenner, 1998);
- offering extended oral presentations of the test materials (Hafner, 2001);
- orally translating the test (Shepard, Taylor, & Betebenner, 1998);
- repeating instructions (Shepard, Taylor, & Betebenner, 1998);
- allowing the students to give their answers orally (Shepard, Taylor, & Betebenner, 1998);
- allowing the students to write their answers in their first language (Shepard, Taylor, & Betebenner, 1998);

- having an adult transcribe part of a student's writing (Shepard, Taylor, & Betebenner, 1998);
- placing the students in a different room (Shepard, Taylor, & Betebenner, 1998);
- testing the students individually (Shepard, Taylor, & Betebenner, 1998);
- having the student sit in the front of the classroom (Shepard, Taylor, & Betebenner, 1998);
- adding more breaks (Shepard, Taylor, & Betebenner, 1998); and
- elongating the testing to more days (Shepard, Taylor, & Betebenner, 1998).

The following is a presentation and discussion of some of the findings from studies investigating testing accommodations for ELL students.

Accommodation vs Modification

Although the preceding list refers to various accommodations, there are some that could be perceived as modifications to the test which can alter the difficulty level or the content being tested. In the preceding list, this could include translating the test, depending on how true to the original the translation is. It could also include adding a glossary or putting the translation of some words in the margin if the purpose of the test is to determine vocabulary skills. If the test is meant to be timed, then extending the time to work on the test could alter the meaning of the student's score. If an adult is providing assistance by transcribing the responses, repeating instructions, orally translating the test, reading the test questions aloud, or offering extended oral presentations, there is a chance that the adult involved may, knowingly or not, give assistance such as stressing a particular word when reading the question aloud. The adult could also reduce the student's score if they do not translate well or miss something when transcribing. All this being said, when used appropriately, these methods which people have used to assist ELL students could still be termed accommodations, not modifications.

Taxonomy

There are two broad classification systems for accommodations found in the previous research. The most commonly used system involves alteration of the presentation method, response method, setting, or timing/scheduling of the test (U.S. Department of Education, 2003). A presentation method accommodation would include such changes as translating the test into the student's first language, making the language less complex, adding visual aids, and allowing the students to use dictionaries. Response method accommodations include allowing the student to write his/her answers in the test booklet, allowing a verbal response, or allowing a response in the student's first language. Setting accommodations include administering the test in small groups or putting students into separate rooms. Timing or scheduling accommodations include allowing more breaks over the test administration, breaking the testing span into multiple sessions, or giving the student extra time to complete the test.

A relatively new system, created by Rivera and Collum (2006), uses two main categories, direct linguistic support and indirect linguistic support. They argued that this taxonomy is more useful in the discussion of ELL accommodations and that the “classical” way was based on the taxonomy of accommodations for students with disabilities. Direct linguistic support includes anything related to the language of the test, such as translating the test, allowing students to respond in their native language, repeating instructions in either language, and simplifying the language used on the test. Indirect linguistic support would include any accommodation that has no direct relation to the language of the test, such as extending the time to take the test, administering the test in a small group setting, adding more breaks to the testing schedule, or having a familiar person administer the test.

Of the states that had a taxonomy in a public accommodations guide, there was a split between whether Rivera and Collum’s (2006) method was used, or whether the “classical” method was used, with slightly more states still using the “classical”

taxonomy. While this alternate taxonomy may become common over time, currently the “classical” way is prevalent among accommodations research and will be used throughout this chapter as a way to organize the previous research on accommodations.

Presentation Accommodations

Dictionaries

Dictionaries can be problematic as an accommodation because there is a level of knowledge of how to use a dictionary needed. In a study by Albus et al (2001), ELL students’ scores on a reading test were positively correlated to their self-reported skill with using a dictionary and those who felt their skills were low had corresponding low scores. Albus et al’s own test found that the ELL students had at least basic dictionary skills, so a simple test on dictionary skills may not be the best method to determine whether ELL students will find using a dictionary helpful. Because the No Child Left Behind legislation (NCLB; 2002) states that the accommodation to be used should be one that students have used in the classroom, it may be that the ELL students will be adept enough on a state assessment for there to be no concern. It may be a concern when researchers are randomly assigning accommodations, however.

Of the researchers who chose to allow students to use dictionaries as an accommodation, there were different approaches in what dictionary would be used. Some researchers specified a particular version of dictionary which they provided (Abedi et al, 2005; Albus et al, 2005). Other researchers allowed students to bring their own dictionaries (Bensoussan, Sim, & Weiss, 1984; Nesi & Meara, 1991). And still other researchers created their own customized dictionaries (Abedi, Lord et al, 2001; Abedi, Courtney, & Leon, 2003).

Two of the studies had results showing that using a dictionary could significantly affect test scores. Abedi, Lord et al (2001) found that a customized English dictionary placed at the end of a test raised ELL students' scores, and did not raise non-ELL

students' scores, providing evidence that this could be a useful accommodation for ELL students. Abedi, Courtney, and Leon (2003) were not able to replicate these findings in a later study using similar methods, however. Albus et al (2005) found that a monolingual English dictionary could improve scores, but only with the ELL students who had an intermediate level of English proficiency.

Nesi and Meara (1991) gave the New Scientist reading test to 83 post-graduate students from countries with languages other than English as their official language. The study was completed at Warwick University in England. Participants were allowed to use English dictionaries, bilingual dictionaries, or no accommodation. Forty-three students brought their own dictionaries to the test, either monolingual or bilingual, and the other forty did not bring any dictionary. No differences were found in scores between groups.

In a second experiment from the same project, Nesi and Meara gave the same test to 65 additional overseas students. This time 34 were given a specific English dictionary, the Oxford Advanced Learner's Dictionary, and 31 were not given a dictionary to use. There were still no differences in performance between groups.

Abedi, Hofstetter et al (2001) gave National Assessment of Educational Progress (NAEP) math items to 8th-graders from Los Angeles, CA, and Long Beach, CA. Some students (146 ELL, 121 non-ELL) were given a glossary as an accommodation, while others (29 ELL, 30 non-ELL students) were given both extra time and a glossary. Neither non-ELL nor ELL students had higher scores in the glossary condition when compared to their counterparts (144 ELL, 130 non-ELL students) in the no accommodation condition. The condition where students had both a glossary and extended time, however, resulted in the best performances for both ELL and non-ELL students.

Abedi, Lord et al (2001) gave NAEP science items to 8th-graders from southern California. Students were given no accommodation (58 ELL, 79 non-ELL), a customized

English dictionary at the end of the test (55 ELL, 82 non-ELL), or Spanish translations of certain words in the margins of the test booklet (70 ELL, 75 non-ELL). Non-ELL students had similar performance across conditions. The ELL students who had translations in the margins performed similarly to those who had no accommodation. The ELL students who had the customized dictionary at the end of the test, though, performed better than the ELL students in the other conditions.

Abedi et al (2005) found that using a standard, commonly-used dictionary had similar results to the Abedi et al (2001) study. The researchers gave 4th- and 6th-graders from Hawaii, California, and Texas NAEP science items. The students were allowed to use an English dictionary (59 ELL, 88 non-ELL students) or a bilingual dictionary (64 ELL students), or were not given any accommodation (62 ELL, 85 non-ELL students). Although a significant difference between using a dictionary and not having any accommodation was found, with those using the dictionary having higher scores, this was just one of many comparisons tested. It might have been prudent to correct for type I error across multiple hypothesis tests. If the researchers had used a Bonferroni correction, they probably would have concluded that the results were actually not significant in either study.

Abedi, Courtney, and Leon (2003) gave math items from the NAEP and TIMSS to 4th-graders and 8th-graders from southern California. Students were given a customized English dictionary (64 ELL, 93 non-ELL 4th-grade students and 86 ELL, 87 non-ELL 8th-grade students) or no accommodation (80 ELL, 98 non-ELL 4th-grade students and 86 ELL, 131 non-ELL 8th-grade students). Both conditions yielded similar results at both grade levels.

Bensoussan, Sim, and Weiss (1984) administered a reading test to 91 undergraduate university students in Israel, all of whom had had several years of English instruction, but whose first language was not English. The students were given an

English dictionary, a bilingual dictionary, or no accommodation. No difference was found in how well the students in the three conditions performed.

The researchers replicated their study two more times, but allowed the students to choose whether they brought a dictionary and what dictionary they would bring. There were 670 and 740 students in these additional experiments. The results were the same as when the researchers had made the condition decision; that there was no difference in scores dependent on the condition to which a student belonged.

Albus et al (2005) gave 202 8th-grade students (69 non-ELL, 133 ELL) items which were designed to be parallel to the Minnesota's Basic Standards Reading Test. This particular sample of ELL students were all Hmong. The researchers provided an English dictionary or no accommodations for the first half of the test; then the same students were given the opposite condition for the second half of the test. The results were mixed, but showed no overall difference between those who were given dictionaries and those who were not.

The researchers found that there was an interaction between the accommodation condition and English proficiency (Albus et al, 2005). For those students who had an intermediate level of English proficiency, the dictionary seemed to be helpful. For those students who had a low level of English proficiency, scores were similar in both conditions. There were instances of the students in the dictionary condition performing differently than expected, with some performing less well than they had in the condition with no accommodation, as well as instances where less proficient students answered certain items right more often than more proficient students.

Computerized Testing

Computerized testing is both a change in the presentation of the test and a change in the way a student responds to test items. So far, there is not much research on how computerized testing might be effective as a testing accommodation. Abedi, Courtney,

and Leon (2003) gave math items from the NAEP and TIMMS to 4th-graders and 8th-graders from southern California. In one condition, students (35 ELL, 44 non-ELL 4th-grade students and 84 ELL, 68 non-ELL 8th-grade students) were given the test items on a computer instead of using a paper form, and students had a pop-up glossary feature available. Other students (80 ELL, 98 non-ELL 4th-grade students and 86 ELL, 131 non-ELL 8th-grade students) were given the same items in a paper form. The researchers found that using a computer for testing instead of a paper-and-pencil format improved ELL students' scores but did not affect non-ELL students' scores. These results are confounded because the computerized version also had a pop-up glossary feature, but there was no condition for a computerized version without the glossary, nor a condition that had a paper version of the glossary as an accommodation. There was another condition in which students were given a customized dictionary with the paper version of the test, though.

Examining the results further, although there was a significant difference between having the computerized version of the test with the pop-up glossary as an accommodation and having no accommodation, there was no significant difference between this accommodation and the customized dictionary condition, or between the customized dictionary condition and having no accommodations. Although this is only speculation, these comparisons seem to indicate that the computerized testing condition not only had some overlap with the customized dictionary condition, but also something unique which made it more successful as an accommodation. Because a glossary and customized dictionary are quite similar, it is possible that the computerized version without the pop-up glossary would also have been successful.

Linguistic Modification

The goal of linguistic modification is to simplify the language of test items without altering the task in order to reduce the lack of English proficiency as a barrier to

being successful on the test (Abedi and Lord, 2001). Although there may be some argument about what makes text easier to process without changing the validity of the item, Abedi and his colleagues, who have done the most research with this accommodation, have created a process for judgment which is described consistently from study to study. Abedi, Lord, and Hofstetter (1998) found that simplifying test items linguistically made some of them easier, but most research on linguistic modification showed that this accommodation did not provide an advantage (Abedi & Lord, 2001; Abedi et al, 2005; Abedi, Hofstetter et al, 2001; Abedi, Lord, & Plummer, 1997; Hofstetter, 2003; Rivera & Stansfield, 2003) or that the ELL students only showed an advantage because the accommodation disadvantaged non-ELL students (Abedi, Hofstetter et al, 2001). Some research suggested the accommodation actually caused a decrease in scores (Rivera & Stansfield, 2003).

Abedi, Lord, and Plummer (1997) gave 1,031 8th-grade students (31% categorized as ELL) from the Los Angeles, CA, area NAEP math items. The researchers created two forms with the same questions, but with some of the items linguistically modified. On one form, the first 10 of the 25 questions were modified, and the second form had the second 10 questions modified. Out of the 1,031 students, 525 received the first form, and 506 received the second form. They found no significant difference between the students' performances in the two conditions.

Abedi and Lord (2001) replicated Abedi, Lord, and Plummer's (1997) work. They presented 1,174 8th-grade students (31% categorized as ELL) from the Los Angeles, CA, area with released items from the NAEP math test, as well as parallel items which had been linguistically modified. Of the 1,174 students, 599 were given the first form, with the first 10 of 25 questions modified, and 575 were given the second form, with the second 10 of the 25 questions modified. There were no significant differences in the performances between the two forms when comparing modified and non-modified questions.

Similar results were obtained by Abedi et al (2005) when they gave 4th- and 6th-graders from Hawaii, California, and Texas NAEP science items. The students either had a test form where some items had been linguistically modified (20 ELL, 23 non-ELL 4th-grade students and 13 ELL 8th-grade students) or were not given any accommodation (62 ELL, 85 non-ELL 4th-grade students and 22 ELL, 33 non-ELL 8th-grade students). Performances were similar across both conditions.

In a similar study, but with slightly different results, Abedi, Hofstetter et al (2001) gave NAEP math items to 8th-graders from Los Angeles, CA, and Long Beach, CA. The 117 non-ELL students who were in the linguistic modification condition had lower scores on average than the 130 non-ELL students who had no accommodation. The ELL students' scores showed no significant difference between the linguistic modification (124 ELL students) and no accommodation conditions (144 ELL students).

Hofstetter (2003) gave NAEP math and reading tests to 8th-graders in Southern California. Students were given a form which either had been translated into Spanish (210 ELL, 24 non-ELL), had some items linguistically modified (228 ELL, 82 non-ELL students) or had no accommodation (238 ELL, 61 non-ELL students). Those at the lowest math levels had higher scores in the linguistic modification condition.

Rivera and Stansfield (2003) gave 11,415 4th- and 6th-graders in Delaware the Delaware Student Testing Program science test. Out of the six forms available, two were not used for this study (form A and form B), two were used with the original English (no accommodation, form C and form D), and the last two forms (form E and form F) had the same questions as forms C and D, respectively, but had some of the items linguistically modified.

The 34 4th-grade ELL students who were in the linguistic modification condition did not perform differently on average than the 38 4th-grade ELL students who had no accommodation for either pair of forms. Non-ELL 4th-grade students, however, showed higher scores on form C, the form with no accommodation (given to 2,856 4th-grade non-

ELL students), than form E, the form with linguistic modification (given to 2,828 4th-grade non-ELL students). There was no difference for the other pair of forms.

The 15 6th-grade ELL students who were in the linguistic modification condition with form E also had lower scores on average than the 22 6th-grade ELL students who had no accommodation on form C. Those who had the other pair of forms, however, showed no difference. Non-ELL 6th-grade students did not show any differences for either pair of forms (2,791 6th-grade non-ELL students were given the linguistic modification forms and 2,831 6th-grade non-ELL students were not given an accommodation). These results are mixed, but do not provide support for linguistic modification as a helpful accommodation for ELL students.

Translated Test

Translating the test into a student's first language is a more complex accommodation than some others for several reasons. First, translating into another language is not straightforward. Second, every new form of the test would have to be translated. Third, there are dozens of potential languages and dialects which would need to have their own translations.

The results of research using translated tests are also not straightforward. The potential value of using a test in another language may depend on whether the student was instructed on that material in that same language (Hofstetter, 2003; Shepard, Taylor, & Betebenner, 1998). Even if students speak Spanish at home, if they are not also being instructed in Spanish at school, they may not do as well on a math test written in Spanish as they would on one written in English (Hofstetter, 2003).

Shepard, Taylor, and Betebenner (1998) collected data from the operational administration of the Rhode Island State Assessment Program. They focused on the math test given to 12,042 4th-graders. Some students, both non-ELL and ELL, had been assigned accommodations by their schools. Other students, even though designated as

ELL, were not given any accommodation. Among the accommodations were three forms of translating the test – written translation of the test (3 general education students, 36 ELL students who had less than 2 years of education in the U.S., 22 ELL students with two or more years of education in the U.S.), oral translation of the test (3 general education students, 26 ELL students who had less than 2 years of education in the U.S., 19 ELL students with two or more years of education in the U.S.), and allowing the students to write their responses in their home language (1 general education student, 15 ELL students who had less than 2 years of education in the U.S., 20 ELL students with two or more years of education in the U.S.). Students, whether non-ELL or ELL, who had not had any accommodations did better than their peers who had an accommodation which involved translating the test.

Hofstetter (2003) gave NAEP math and reading tests to 8th-graders in Southern California. Students were given a form which had been translated into Spanish (210 ELL, 27 non-ELL), had some items linguistically modified (228 ELL, 83 non-ELL students), or had no accommodation (238 ELL, 63 non-ELL). All of the students who were given the form in Spanish knew Spanish. Furthermore, some of the students had been participating in a program in which the language of instruction was Spanish. Students did best in whichever language had been their language of instruction.

Read Aloud

There are mixed results for reading test questions aloud, with Shepard, Taylor, and Betebenner (1998) finding that this accommodation can be helpful to some students, and Castellon-Wellington (2000) not finding any significant difference between having the accommodation and having no accommodation.

Castellon-Wellington (2000) gave the ITBS Social Studies Test to 106 7th-grade students from Burbank, CA. Some students were given no accommodation, while others

had the test read aloud to them. They found no difference between the two groups' performances.

Shepard, Taylor, and Betebenner (1998) collected data from the operational administration of the Rhode Island State Assessment Program. They focused on the math test given to 12,042 4th-graders. Some students, both non-ELL and ELL, had been assigned accommodations by their schools. Other students, even though designated as ELL, were not given any accommodation. Reading the test aloud was among the accommodations given to students. This accommodation was found to be associated with higher scores for the ELL students who had been in an ELL program two or more years, but not for ELL students who had been in an ELL program less than two years or for non-ELL students.

Repeating Instructions

Another accommodation given to ELLs in the study by Shepard, Taylor, and Betebenner (1998) was repeating instructions. This accommodation was found to be associated with higher scores for the ELL students who had been in an ELL program two or more years, but not to ELL students who had been in an ELL program less than two years or to non-ELL students.

Extended Oral Presentation

Hafner (2001) gave 292 4th-grade and 159 7th-grade students from California and New Mexico the TerraNova math test. Students were given extended oral presentation and extra time, or given no accommodations. The extended oral presentation included simplifying test directions, re-reading directions, providing examples, and reading in the student's first language. Scores in the extended time with extended oral presentation were similar to those in the no accommodation condition.

Response Accommodations

Shepard, Taylor, and Betebenner (1998) found that students who had been in an ELL program for two or more years were helped by having an adult transcribe part of their writing and by giving their responses orally. Non-ELL and ELL students who had been in an ELL program for less than two years also had higher scores if allowed to give their responses orally.

Setting Accommodations

Shepard, Taylor, and Betebenner (1998) also examined some setting accommodations – testing in another room (such as special education or ELL classrooms), testing with a small group, testing individually, and seating the student in front of the classroom. They found that small group testing and seating the student in front of the classroom did not help students perform better than their peers who did not receive any accommodations. Only ELL students who had been in an ELL program two or more years were helped by being in another room during testing. Individual testing helped both non-ELL students and ELL students who had been in an ELL program two or more years. None of these four accommodations appeared to help ELL students who had been in an ELL program less than two years.

Abedi, Courtney, and Leon (2003) gave math items from the NAEP and TIMSS to 4th-graders and 8th-graders from southern California. One of the accommodations used in the study was small group testing. They did not find any significant difference between scores for students who had this accommodation (11 ELL, 9 non-ELL) and students who did not (80 ELL, 98 non-ELL).

Timing/Scheduling Accommodations

Extended Time

Abedi, Courtney, and Leon (2003) found that giving extra time improved ELL students' scores while it did not affect non-ELL students' scores. In opposition, Castellon-Wellington (2000) found no significant difference between having extended time as an accommodation and having no accommodation at all. Most studies, however, found that the extra time helped both ELL and non-ELL groups (Abedi, Hofstetter et al, 2001; Hafner, 2001). Shepard, Taylor, and Betebenner (1998) found that different forms of extended time produced different outcomes for non-ELL students, ELL students who had less than two years of English instruction, and ELL students who had two years or more of English instruction.

Shepard, Taylor, and Betebenner (1998) collected data from the operational administration of the Rhode Island State Assessment Program. They focused on the math test given to 12,042 4th-graders. Some students, both non-ELL and ELL, had been assigned accommodations by their schools. Other students, even though designated as ELL, were not given any accommodation. Among these accommodations were three forms of extended time - extending the time allowed for completing the test in one session, more frequent breaks during testing, and extending testing sessions over several days.

Extending the session time, given to 424 general education students, 24 ELL students with less than 2 years of education in the U.S., and 138 ELL students with two or more years of education in the U.S., was found to be helpful to non-ELL students and ELL students who had two or more years of education in the U.S., but not helpful to ELL students who had less than two years of education in the U.S.. Only the ELL students who had been in school in the U.S. two or more years were helped by having more frequent breaks during testing. The reverse pattern was found for extending the testing

sessions over several days, with the ELL students who had been in the U.S. fewer than two years having higher scores than their counterparts who had not had any accommodation.

Castellon-Wellington (2000) gave the ITBS Social Studies Test to 106 7th-grade ELL students from Burbank, CA. Students were first given the test with no accommodation. They were then asked if they would rather be given extra time on the test or have the items read aloud to them. The researchers gave an equivalent form of the test a week later, giving the students the accommodation they preferred, giving them the opposite accommodation from their preferred accommodation, or giving an accommodation at random. There were no significant differences in the scores between groups.

Hafner (2001) gave 4th-grade students from California and New Mexico the TerraNova math test. Students had extra time to complete the test (40 ELL, 60 non-ELL), extended oral presentation and extra time (40 ELL, 60 non-ELL), or no accommodation (40 ELL, 60 non-ELL). For both ELL and non-ELL students, the extended oral presentation included simplifying test directions, re-reading directions, providing examples, and reading in the student's first language. The extended time condition had higher scores than the other two conditions. Scores in the extended time with extended oral presentation were similar to the no accommodation condition.

Abedi, Hofstetter et al (2001) gave NAEP math items to 8th-graders from Los Angeles, CA, and Long Beach, CA. Some students were given extra time as an accommodation (30 ELL, 25 non-ELL), while others were given both extra time and a glossary (29 ELL, 30 non-ELL). Both non-ELL and ELL students had higher scores in the extended time condition when compared to their counterparts in the no accommodation condition. The condition where students had both a glossary and extended time resulted in the best performances for both ELL and non-ELL students.

Abedi, Courtney, and Leon (2003) obtained similar results when they gave math items from the NAEP and TIMMS to 4th-graders from southern California. Students either had extra time to complete the test (89 ELL, 84 non-ELL) or no accommodation (80 ELL, 98 non-ELL). While the non-ELL students had similar scores in both conditions, ELL students in the extended time condition had higher scores than those in the no accommodation condition.

Other Research

Some researchers did not analyze their results in a way that would allow for comparing the accommodations which were used to each other, or to having no accommodation. Kopriva et al (2007), for example, compared the recommended accommodations to randomly-assigned accommodations. The authors did not discuss how effective the accommodations used in their study (using a picture dictionary, adding a bilingual glossary, and having test questions read aloud) were when compared to each other. Because all participants were given a computerized version of the test, there is also no way to extricate the effects of using a computer as an accommodation, and there could be interactions between using the computer and the other accommodations. The results offered enough information to determine that the randomly-selected accommodations were not better than having no accommodation, and that the recommended accommodations were better than the randomly-selected accommodation or no accommodation, but no information was given that would detail how each accommodation performed when it was used as recommended.

Liu et al (1999) also had another focus besides determining how effective an accommodation could be. The accommodation used in their study was adding a set of instructions in the student's first language to a reading test. They had only nine students in their sample, so they were limited in their analyses. They analyzed how much the accommodation was used and how well it was liked. The usage of the two sets of

instructions was varied, but the students who had been in the U.S. longer seemed to prefer only using the English instructions.

Important Considerations for Deciding on Accommodations

Because ELL students are not all the same, it is likely that accommodations will not work the same for all (Butler & Stevens, 1997; Stevens, Butler, & Castellon-Wellington, 2000). A student's ability to learn is affected by his/her home life, individual characteristics, educational background, community, and language (Butler & Stevens, 1997). The accommodation choice should be based on the background of the student, and the validity, effectiveness, and feasibility of the choice (Abedi, Hofstetter, & Lord, 2004). Students at different levels of English proficiency may use accommodations with different levels of effectiveness (Abedi, Hofstetter, & Lord, 2004; Albus et al, 2005; Shepard, Taylor, & Betebenner, 1998). The Center for Research on Evaluation, Standards, and Student Testing (CRESST) suggests using academic English proficiency, prior formal schooling, and length of time in the United States as the most practical information to focus on when selecting an appropriate accommodation for an individual (Butler & Stevens, 1997).

There is a difference between ELL students who have already had some instruction in their own language and those who have not (Collier, 1987; Fairbairn & Fox, 2009). The students who have had instruction in their home language perform better on academic tests and learn English more quickly than those who have not (Collier, 1987). Students do better on tests written in the same language used in instruction (Abedi, Lord, & Hofstetter, 1998; Hofstetter, 2003). Accommodations which are deemed appropriate for a student should improve tests scores, whereas giving an accommodation which is not appropriate for the student should result in the same effect as having no accommodation (Kopriva et al, 2007).

Ethnicities in Previous Research

In most studies, the majority of ELL participants have been Hispanic (Abedi & Lord, 2001; Abedi, Hofstetter et al, 2001; Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Plummer, 1997; Castellon-Wellington, 2000; Hafner, 2001; Shepard, Taylor, & Betebenner, 1998), and some studies even limited their ELL subject pool to Hispanic students (Hofstetter, 2003; Kopriva et al, 2007; Liu et al, 1999). Another study limited the ELL subject pool to students of Hmong descent (Albus et al, 2005). Many other ethnicities have been represented, including Korean (Abedi & Lord, 2001; Abedi et al, 2005; Abedi, Lord, & Plummer, 1997; Castellon-Wellington, 2000; Nesi & Meara, 1991), Chinese (Abedi & Lord, 2001; Abedi et al, 2005; Abedi, Lord, & Plummer, 1997; Nesi & Meara, 1991), Filipino (Abedi & Lord, 2001; Abedi et al, 2005; Abedi, Lord, & Plummer, 1997; Castellon-Wellington, 2000), Farsi (Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997), Hebrew (Bensoussan, Sim, Weiss, 1984), Arabic (Bensoussan, Sim, Weiss, 1984; Castellon-Wellington, 2000), French (Bensoussan, Sim, Weiss, 1984; Nesi & Meara, 1991), Russian (Bensoussan, Sim, Weiss, 1984), Vietnamese (Castellon-Wellington, 2000), Armenian (Castellon-Wellington, 2000), Japanese (Nesi & Meara, 1991), Thai (Nesi & Meara, 1991), Polish (Nesi & Meara, 1991), and Turkish (Nesi & Meara, 1991). None of these studies used ethnicity as a variable in their analyses, though, so it is not known if that would play a role in specifying an accommodation.

Grade Levels in Previous Research

Most studies have concentrated on 3rd- through 8th- grade students (Abedi & Lord, 2001; Abedi, Hofstetter et al, 2001; Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Plummer, 1997; Albus et al, 2005; Castellon-Wellington, 2000; Hafner, 2001; Hofstetter, 2003; Kopriva et al, 2007; Liu et al, 1999; Rivera & Stansfield, 2003; Shepard, Taylor, & Betebenner, 1998). Only two studies have

used older students, including freshmen college students (Bensoussan, Sim, Weiss, 1984) and post-graduate students (Nesi & Meara, 1991). No studies have used students below 3rd grade. None of the studies which used students from more than one grade level found a difference in the effect of accommodations for the two grade levels (Abedi, Courtney, & Leon, 2003; Abedi et al, 2005; Rivera and Stansfield, 2003; Kopriva et al, 2007).

Assessments Used in Previous Research

Most of the previous research has used the NAEP (Abedi, Courtney, & Leon, 2003; Abedi & Lord, 2001; Abedi, Hofstetter et al, 2001; Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Lord, & Plummer, 1997; Hofstetter, 2003). Other assessments which have been used to study accommodations for ELL students are the Trends in International Mathematics and Science Study (TIMSS) (Abedi, Courtney, & Leon, 2003), Minnesota Basic Standards (MBS) (Albus et al, 2005; Liu et al, 1999), Iowa Tests of Basic Skills (ITBS) (Castellon-Wellington, 2000), TerraNova (Hafner, 2001), Delaware Student Testing Program (DSTP) (Rivera & Stansfield, 2003), Rhode Island State Assessment Program (RISAP) (Shepard, Taylor, & Betebenner, 1998), and a researcher-developed test based on the South Carolina state standards similar to the state standardized achievement test (Kopriva et al, 2007).

Shepard, Taylor, & Betebenner (1998) used the math portion of the RISAP. Items covered tasks that included matching a story to data on a graph, estimation, multiplication and division applications, representing numbers with base 10 stickers, and representing tangrams with numbers. All of the items were constructed-response and required students to explain their answers. There were 10 multi-part items given over two days, with a one-hour session each day. If students were still working on the test at the end of the session, then they were given 10 more minutes to work.

Castellon-Wellington (2000) used the ITBS social studies test. Participants were given 44 multiple choice questions which covered history, geography, political science,

economics, sociology, and anthropology. The participants were given 30 minutes to complete the test.

Abedi, Lord, & Plummer (1997) chose 20 NAEP-released math items. These items covered addition, subtraction, multiplication, division, calculating rate/time/distance, fractions, proportions, measurements and weights, geometry, pre-algebra, algebra, and reading graphs and tables. Participants were allowed one hour to complete their test. For the version of these items that had been modified to make the English simpler, there were 16 multiple choice and 4 constructed response items. The original version had 15 multiple choice and 5 constructed response items.

Abedi and Lord (2001) also chose 20 NAEP-released math items. It is not clear how much time was given to participants to complete these items, nor how many of these items were constructed response or multiple choice.

Abedi, Hofstetter et al (2001) also chose NAEP-released math items. They selected 35 items, 24 of which were multiple choice and 11 of which were constructed response. They allowed 45 minutes for participants to complete the items. Hofstetter (2003) used the same format for her test.

Abedi, Lord et al (2001) chose 20 NAEP-released science items, all of them multiple choice. The items covered earth, physical, and life science concepts. The authors did not report how much time was allowed to complete their test.

Abedi et al (2005) also selected items from the released NAEP science test. They chose 19 items, 11 of them multiple choice and 8 constructed response, for the 4th graders. They chose 30 items, 24 of them multiple choice and 6 constructed response, for the 8th-graders. Both grade levels were allowed 45 minutes to complete the test.

Abedi, Courtney, and Leon (2003) used released NAEP and TIMSS math items. It is unclear how many items were selected, and, although the authors indicated that they used some multiple choice items and some constructed response items, they did not make clear how many of each were used, or in what proportion. They allowed 4th-grade

students to have 40 minutes and 8th-grade students to have 30 minutes, with the extended time condition receiving up to 60 minutes.

There were some studies which created their own test items instead of using ones that already existed and were from known tests. The researchers still attempted to mimic known tests, however. Both Albus et al (2005) and Liu et al (1999) created two reading passages for each form, and had two forms. The test items were 10 multiple choice questions for each passage. The questions included literal or inferential comprehension questions. The passages and associated questions were created to imitate those found on the Minnesota Basic Standards reading test. Participants were given up to two hours to complete the test.

Rivera & Stansfield (2003) also created test items that imitated their state exam, the Delaware Student Testing Program science test in this instance. Because this was an operational test, they were allowed to include additional field test items. Out of the 50 items each participant answered, 10 were field test items. Of the 50 items, 32 were multiple choice and 18 were constructed response. Of the field test items, 6 were multiple choice and 4 were constructed response. Thus, the field test was created to be a miniature version of the regular test. The authors did not indicate how much time was available for taking the test.

Kopriva et al (2007) used the South Carolina Department of Education mathematics standards for 3rd-grade students as their base, aiming to have similar questions as the South Carolina state achievement test. They created 33 items, which included 30 multiple choice items and 3 constructed response items. Only the multiple choice items were used in the analysis for the study, though. It is not clear how much time was given to participants to complete the test. Although some studies used a computerized version of the test as an accommodation, this was the only study to use it as the only method of presenting the test.

Locations of Previous Research

As mentioned near the end of the first chapter, most studies on the effectiveness of testing accommodations for ELL students were conducted in California (Abedi & Lord, 2001; Abedi, Hofstetter et al, 2001; Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Plummer, 1997; Castellon-Wellington, 2000; Hafner, 2001; Hofstetter, 2003). Other states which have contributed to research findings are Hawaii (Abedi et al, 2005), Texas (Abedi et al, 2005), Minnesota (Albus et al, 2005; Liu et al, 1999), New Mexico (Hafner, 2001), Delaware (Rivera & Stansfield, 2003), South Carolina (Kopriva et al, 2007), and Rhode Island (Shepard, Taylor, & Betebenner, 1998). There has also been some pertinent research done in England (Nesi & Meara, 1991) and Israel (Bensoussan, Sim, Weiss, 1984).

Because this current research includes data from throughout the United States, it is important to demonstrate how states which have had related research are different from each other and the nation as a whole. Tables 1 through 3 provide information for comparing the eight states where research has previously taken place in terms of ethnicity, poverty, and ELL students.

Ethnicities

The National Center for Education Statistics (NCES), a part of the U.S. Department of Education, annually collects educational data on a national level from state education agencies and compiles it into the Common Core of Data. Table 1 (derived from NCES, 2009) describes the percentages of several ethnic groups in the states' schools. The American Indian/Alaska Native group has the lowest percentages on the table. New Mexico is the only one of these states to have over 5% of its student population listed as American Indian/Alaska Native, with 11% of the students in this group. Minnesota, with its 2%, has slightly more than the national average of 1%. The other states have less than 1% of their students listed as American Indian/Alaska Native.

Almost three-fourths of Hawaii's students are Asian/Pacific Islanders. California also has a good-sized population of Asian/Pacific Islanders, but they make up only 12% of the student population compared to 73% in Hawaii. Minnesota is the only other state to have over 5% of their student population listed as Asian/Pacific Islanders, and they only have 6%. The country as a whole has just under 5% of students listed as Asian/Pacific Islanders. Delaware, Rhode Island, and Texas all have just over 3% of their student population made up of Asian/Pacific Islanders. South Carolina and New Mexico have the smallest percentages of this ethnic group, with just over 1%.

Table 1. Percentage of Students in Ethnic Groups by State

State	American Indian/Alaska Native	Asian/Pacific Islander	Hispanic	Black, non-Hispanic	White, non-Hispanic
California	0.73%	11.61%	48.47%	7.19%	27.55%
Delaware	0.36%	3.37%	10.95%	33.24%	52.08%
Hawaii	0.62%	72.95%	4.65%	2.34%	19.46%
Minnesota	2.17%	6.19%	6.43%	9.56%	75.65%
New Mexico	11.04%	1.36%	56.07%	2.63%	28.89%
Rhode Island	0.73%	3.18%	18.51%	9.01%	68.57%
South Carolina	0.37%	1.56%	5.42%	38.65%	53.49%
Texas	0.35%	3.57%	47.89%	14.16%	34.02%
United States	1.19%	4.92%	21.23%	16.76%	54.25%

The Hispanic ethnic group has a large range across these states. At the low end for the Hispanic group are Hawaii, South Carolina, and Minnesota, each close to 5%. Delaware has about 11%, and Rhode Island about 19%. The national average is 21%.

The states with the highest percentages of Hispanic students are California and Texas, at 48%, and New Mexico at 56%.

The lowest percentages of Black, non-Hispanic students are found in Hawaii and New Mexico, around 2%. Next, from 7-10%, are California, Rhode Island, and Minnesota. Texas has approximately 14%, just under the national average of 17%. Delaware, with 33%, and South Carolina, with 39%, have the largest percentages of Black, non-Hispanic students.

Of the eight states considered, Hawaii has the lowest percentage of White, non-Hispanic students, close to one-fifth of its students. Three of the states, California, New Mexico, and Texas, also have lower numbers, close to one-third of their students listed as White, non-Hispanic. South Carolina and Delaware are most like the national average of 54% for this group. Rhode Island, with 69%, and Minnesota, with 76%, have the largest groups of White, non-Hispanic students.

Poverty

Table 2 (derived from NCES, 2009) contains data on poverty with percentages of students eligible for free or reduced-price lunch in each state studied. New Mexico, South Carolina, California, and Texas, all have more students eligible for free or reduced-price lunch than the national average. About half of the students in California and South Carolina, and more than half of the students in New Mexico, are eligible. Delaware, Hawaii, and Rhode Island are closer to the national average. Minnesota has the lowest percentage of students eligible for free or reduced-price lunch among these states.

Table 2. Percentage of Students Eligible for Free or Reduced-Price Lunch by State

State	Percentage Eligible for Lunch Programs
California	51.56%
Delaware	39.05%
Hawaii	41.73%
Minnesota	32.39%
New Mexico	61.40%
Rhode Island	39.34%
South Carolina	52.46%
Texas	48.76%
United States	43.64%

Percentage of ELL Students

Table 3 (derived from NCES, 2008, and NCES, 2009), shows percentages of students involved in ELL programs for each state studied. California and New Mexico far outpace the national average, with Texas not far behind. Hawaii and Minnesota are nearer to the national average. Rhode Island, Delaware, and South Carolina are on the low end of percentages of students receiving ELL services. Not surprisingly, because Hispanic students make up the majority of the ELL student population (NCES, 2009), these data are somewhat consistent with the data on percentages of Hispanic students in each state, with California, New Mexico, and Texas having more Hispanic students and also having more students receiving ELL services.

When ethnicity, poverty, and percentage of ELL students are compared across these eight states and the national average, the conclusion is that there is quite a bit of variability; each state has a unique combination of students. These data demonstrate that there can be major differences between locales which make it potentially unsound to

make generalizations from studies which do not take into account the possible effects of their participants' individual characteristics.

Table 3. Percentage of Students Receiving ELL Services by State

State	Percentage Receiving ELL Services
California	23.96%
Delaware	5.73%
Hawaii	10.34%
Minnesota	7.35%
New Mexico	18.60% *
Rhode Island	4.84% *
South Carolina	4.38%
Texas	15.12%
United States	8.84%

*These percentages come from NCES data for 2007-2008 because the data were not reported for 2008-2009

Operational Testing

In an experimental testing environment, there may be less authenticity than an operational administration in important variables such as motivation, using the entire test, and testing the students in realistic conditions. The only previous studies that used an operational test setting were the 2003 study in Delaware (Rivera & Stansfield) and the 1998 study in Rhode Island (Shepard, Taylor, & Betebenner).

Only one study, Shepard, Taylor, and Betebenner (1998), did not try to introduce any accommodations. Instead, the accommodations were assigned by school personnel. This, and the fact that this was an operational test instead of an experiment, may allow

the results to be more generalizable than other studies. This study was done before NCLB came into being, however, which means the environment surrounding the test, and educating the students in general, could be different. Before NCLB, state assessments were mainly used as a way to determine how well individual students were performing and whether they needed some sort of remediation or advancement. Now, however, assessments are also used as a way to judge schools, which puts much more pressure on schools to elevate student scores, creating a different atmosphere and motivation surrounding the testing setting. Although NCLB may change or be eliminated in the near future, it still has had an important impact on schools and standardized testing, and the current study resides in a more relevant educational landscape.

The test which the researchers used, the Rhode Island State Assessment Program, was a statewide standardized test. The researchers chose to focus on the 4th-grade math test administered in the spring of 1997 for their analysis. The ELL students were given a wide array of accommodations.

Out of the 12,042 students expected to take the test, 734 were considered to be ELL students. Of these, 94% participated in the testing. This participation rate was lower than the rate (98%) for students not listed as in an ELL or special education program.

The researchers compared students who had been in an ELL program less than two years to those who had been in an ELL program two or more years. In addition, they compared both groups to those in the general education path. They investigated the variability of accommodation usage, and how having an accommodation interacted with various groups of students.

The most widely used accommodations were oral reading of the test, repeating directions, testing in a special classroom or small group, and extended time. In general, those who received accommodations did better than those who did not, whether the students were ELL students or general education students. The least effective

accommodation overall was having the student sit in the front of the room during testing, but there were only a few students who were given that accommodation, and they had extremely low scores, so it is possible that there were issues besides ELL status being managed.

The researchers also found that different accommodations worked better with some subgroups than others. ELL students, both those with fewer than two years in an ELL program and those with more than two years in a program, have some similar, but not always the same, results related to how well different accommodations helped. ELL students who had been in a program less than two years did best when they were allowed to give their responses orally, and worst when using a test that had been translated into Spanish. ELL students who had more than two years in a program did best when an adult transcribed a portion of their writing, and worst when using a test that had been translated into Spanish.

Another important finding was that schools tended to only give accommodations if they had large numbers of ELL students. Those schools with fewer ELL students often did not provide accommodations.

Summary

This chapter reviewed the research designed to explore the effectiveness of testing accommodations for ELL students. A summary of the research can be found in Table 4. Among these accommodations, the current study includes the use of a word-to-word dictionary, extended time, small group testing (allowing the ELL teacher to administer the test), repeating instructions, and reading items aloud.

There was no clear answer for which accommodation is the best. Although this study does not resolve this question either, it includes an examination of variables which future researchers might consider including in their analyses because they could interact with the effect of the accommodation. The current study strove to describe testing

accommodations in terms of who is receiving them and how well ELL students receiving accommodations perform on an achievement test battery. The methods for achieving this are explained in Chapter Three.

Table 4. Summary of Previous Research

Study	Grade	Ethnicity	Language	Location	Test	Operational	Content	Accommodation	Result
Abedi, Courtney, & Leon, 2003	4th, 8th	Mostly Hispanic	Spanish	Southern California	NAEP, TIMSS	No	Math	Computer testing with a pop-up glossary, extended time, a customized dictionary, small-group testing	No differences for non-ELL students. ELL students did better with computer testing and extended time than no accommodation or dictionary.
Abedi, Courtney, Mirocha, Leon, & Goldberg 2005	4th, 8th	Spanish, Chinese, Filipino, Korean	Spanish, Chinese, Filipino, Korean	Hawaii, California, Texas	NAEP	No	Science	Dictionary, bilingual dictionary, linguistic modification	No difference.
Abedi, Hofstetter Baker, & Lord 2001	8th		Many Spanish	Los Angeles, Long Beach, California	NAEP	No	Math	Linguistic modification, glossary in margins, extended time, both glossary and extended time	Scores were lower for non-ELL students with linguistic modification. Glossary made no difference. Extended time helped both ELLs and non-ELLs. Glossary plus extended time helped both groups even more.
Abedi & Lord, 2001	8th	35% Latino, 26% White, 19% Black, 16% Asian, 4% other	Spanish, Korean, Farsi, Chinese, Filipino	Los Angeles	NAEP	No	Math	Linguistic modification	No difference.
Abedi, Lord, and Hofstetter, 1998	8th				NAEP	No	Math	Linguistic modification	Many of the items showed improved performance after they were simplified linguistically
Abedi, Lord Kim, & Miyoshi, 2001	8th	Hispanic, White, Asian, Black, American Indian		Southern California	NAEP	No	Science	Spanish translations of certain words in the margins, customized English dictionary at end of test	ELL students had higher scores with the customized dictionary accommodation. Glossary made no difference. Non-ELLs showed no difference.

Table 4. Continued

Study	Grade	Ethnicity	Language	Location	Test	Operational	Content	Accommodation	Result
Abedi, Lord, & Plummer, 1997	8th	Mostly Hispanic	Spanish, Korean, Farsi, Chinese, Filipino	Los Angeles, California	NAEP	No	Math	Linguistic modification	No difference.
Albus, Thurlow, Liu, & Bielinski 2005	8th	Hmong	Hmong	Twin Cities, Minnesota	MBS (MN state assessment parallel designed)	No	Reading	Dictionary	No difference.
Bensoussan, Sim, Weiss, 1984	University	Mostly Hebrew	Hebrew, Arabic, French, Spanish, Russian	Israel		No	Reading	English dictionary, bilingual dictionary	No difference.
Castellon-Wellington, 2000	7th	Mostly Hispanic, Armenian	Spanish, Armenian, Korean, Filipino, Arabic, Vietnamese	Burbank, California	ITBS	No	Social Studies	Extended time, reading items aloud; preferred vs random	No difference.
Garcia, 1991	5th, 6th	Hispanic	Spanish	Urbana, Illinois	CAT (California Achievement Test) and CTBS (California Test of Basic Skills)	No	Reading	Extended time	Both ELL and non-ELL scores were higher.
Hafner, 2001	4th	Mostly Hispanic	Spanish	California, New Mexico	TerraNova	No	Math	Extended time with and without extended oral presentation	Both ELLs and non-ELLs had higher scores with extended time.

Table 4. Continued

Study	Grade	Ethnicity	Language	Location	Test	Operational	Content	Accommodation	Result
Hofstetter, 2003	8th	Hispanic	Spanish	Southern California	NAEP	No	Math, Reading	Linguistic modification, Spanish translation	Students did best in language of instruction. Those at the lowest math levels had higher scores with linguistic modification.
Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007	3rd, 4th	Hispanic	Spanish	South Carolina	Based on South Carolina Department of Education mathematics standards judged to be consistent in language and load to local standardized achievement tests	No	Math	Picture dictionary, bilingual glossary, oral reading of test items in English, both oral reading and picture dictionary, both oral reading and bilingual glossary, both picture dictionary and bilingual glossary, or oral reading, bilingual glossary, and picture dictionary	Students with appropriate accommodations had higher scores than either no accommodations or random accommodations. Random and no accommodations had the same effect.
Liu, Anderson, Swierzbin, & Thurlow, 1999	7th	Hispanic	Spanish	Minnesota	MBS (MN state assessment parallel designed)	No	Reading	Side by side bilingual items	Inconsistent results.
Nesi & Meara, 1991	Post Grad	Overseas	Japanese, French, Turkish, Chinese, Polish, Korean, Thai	England	New Scientist	No	Reading	English dictionary, bilingual dictionary	No differences found.

Table 4. Continued

Study	Grade	Ethnicity	Language	Location	Test	Operational	Content	Accommodation	Result
Rivera & Stansfield, 2003	4th, 6th			Delaware	DSTP (Delaware state assessment)	Yes	Science	Linguistic modification	No difference, except 6th grade showed students with the accommodations had lower scores.
Shepard, Taylor, & Betebenner, 1998	4th	Mostly Hispanic	Spanish	Rhode Island	Rhode Island state assessment	Yes	Math	Extended time, small group administration, translated test, reading test items aloud, oral translation, repeating instructions, oral response, written response in native language, transcribing, placing in a separate room, individual testing, sitting in front of classroom, adding more breaks, increasing time span to more days	Interaction between length in ELL program and accommodation. For both ELL and non-ELL, using tests translated into Spanish gave worst scores. For ELL students who had been in program longer than 2 years, best was having someone transcribe written portions for them. For those in an ELL program less than 2 years, best was giving response orally.

CHAPTER THREE – METHODS

The purpose of this study was to investigate testing accommodations being given to English Language Learners (ELLs) on a standardized achievement test as a result of administrative practices of the schools participating in the study. The research questions that guide the analyses are:

- 1) What is the relationship between socioeconomic status, migrant status, ethnicity, gender, grade level, home language, and which accommodations the ELL students receive?
- 2) To what extent do Math, Science, and Reading scores vary among ELL students receiving different types of accommodations?
- 3) How does the gap between ELL and non-ELL performance in a norming study compare to low stakes (NAEP) and high stakes (NCLB state assessments) ELL achievement gaps?

In order to answer these questions, migrant status, home language, ethnicity, SES, gender, and grade level are described for ELL students with different types of accommodations. Additionally, Math, Science, and Reading scores are reported for each of the accommodation groups. Statistical analyses are included when appropriate. More information about the procedure and why migrant status, home language, ethnicity, SES, gender, and grade level were chosen as important variables follow.

Sample and Procedure

Data for this study came from an achievement battery's 2010 national standardization. The data for the norming study came from a non-weighted sampling of United States schools. Care was taken to gather data from all geographic regions, public and private schools, and urban and rural areas. It is important to recognize that the unweighted results presented in this study do not generalize to the characteristics of the

weighted sampled used to develop norms and derived scores reported for the assessment. Some groups were oversampled in order to acquire reliable data.

Besides the test items, a form, which can be found in Appendix A, that had variables that have been identified by previous research as important was completed. Students reported home language, gender, and ethnicity, while ELL status, free or reduced-price lunch eligibility, migrant status, and any accommodations used on the test were reported by the test administrator. Other variables that have been identified as important, but are not available to this research study, include parental education level, length of residency in the United States, English proficiency, history of schooling in a native language, and how many times the student has changed schools.

The data were kept in a protected folder to which only the researcher had access. To insure anonymity, no student or school names, or other identifying information were included in the data set.

Participants

There were 33,226 participants in the study, a subset of the data collected by the assessment program. The participants were in grades 3-8, with 9.0% in Grade 3, 20.3% in Grade 4, 20.3% in Grade 5, 19.9% in Grade 6, 15.7% in Grade 7, and 14.9% in Grade 8. Females comprised 49.7% of the sample.

Of the available ethnicity options, allowing students to select as many as they wanted, 5.4% chose American Indian/Alaska Native, 4.3% chose Asian, 1.0% chose Native Hawaiian/Other Pacific Islander, 18.0% chose Black/African American, and 67.6% chose White. These values do not add to 100% because 9.0% of students did not select an ethnicity, while 4.8% selected more than one.

Students were also asked to select between Hispanic/Latino and non-Hispanic/Latino. Twenty-two percent of the sample chose Hispanic/Latino and 60.0% chose non-Hispanic/Latino. Because 0.2% of the students selected both and 18.1% of

students selected neither option, these do not add to 100%. Although students were not supposed to select both options, it is a plausible scenario for mixed-race students, so the data were not removed. Of the 72 students who selected both, only 3 were ELL students, one of whom received the extended time accommodation.

Only 0.3% of the sample was identified as migrants. There were 18.1% who were identified by the test administrators as being in the free or reduced-price lunch program (FRPL). For home language, 77.0% chose English, 10.7% chose Spanish, and 3.1% chose Other. These percentages do not add to 100% because not all students made a selection for this question.

There were 3.8% who were identified by test administrators as being in ELL programs. This number may have actually been higher, however, because there was evidence that there were ELL students who were not properly identified. Although test administrators had marked 167 students as receiving a testing accommodation reserved for ELL students, only 88 of those had also been marked as ELL. For the purposes of this study, the remaining 79 were included as ELL for the analyses. Of the 167 students who had been reported as receiving an accommodation, 51.5% were marked as having been given extended time, 11.4% were allowed to use a dictionary, 1.2% had items read aloud to them, 21.0% had instructions repeated, and 16.2% had the test administered by an ELL teacher. These do not add to 100% because there was one student who had more than one accommodation.

Measures

The achievement test battery is used as a No Child Left Behind (NCLB; 2002) statewide assessment of students from 3rd to 8th grade, as well as a diagnostic tool in many other states. The tests cover reading, written expression, math, science, and social studies, but the areas of interest in this research project were reading, math, and science, because they are the subject areas used in most of the previous research on

accommodations (Abedi & Lord, 2001; Abedi, Hofstetter et al, 2001; Abedi, Lord et al, 2001; Abedi et al, 2005; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Plummer, 1997; Hafner, 2001; Hofstetter, 2003; Rivera & Stansfield, 2003; Shepard, Taylor, & Betebenner, 1998) and are required under NCLB. Other studies have also found reading scores to be significantly positively correlated with other test scores (Castellon-Wellington, 2000; Hafner, 2001; Hofstetter, 2003; Kopriva et al, 2007), and have used reading scores as a control variable, finding it to be helpful in investigating relationships between other variables (Abedi et al, 2005; Abedi, Courtney, & Leon, 2003). Therefore, this study included Reading, Math, and Science test scores (national percentile ranks) to allow for comparisons between groups.

The overall purposes of the achievement test battery are to identify strengths and weaknesses, inform instruction, monitor growth, determine college readiness, measure the mastery of core learning standards, identify students who may benefit from intensive intervention, inform placement decisions, enable making comparisons among students and groups, and evaluate educational programs. The Reading test includes literary and informational passages and has questions that focus on interpreting, analyzing, and extending information. The Math test covers number sense, operations, algebraic patterns and connections, data analysis, probability, statistics, geometry, and measurement. The Science test covers the methods and processes used in scientific inquiry, and knowledge of life science, earth and space science, and physical science.

This study focused on the levels of the test which align with grades 3-8 because students below grade 3 are not required to take the tests for NCLB reporting. Percentile ranks from the sample were used to facilitate comparisons across grades and to enable straightforward interpretations of test performance. According to test specifications, students are given varying numbers of items for a complete test based on their testing level, as described in Table 5. The entire battery of tests takes 2 hours and 35 minutes for grades 3-8.

Table 5. Number of Test Items per Grade for Each Content Area

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Reading	41	42	43	44	45	46
Math	50	55	60	65	70	75
Science	30	34	37	39	41	43

Rationale for Chosen Study Variables

As noted earlier, there are several variables which are of interest in investigating testing accommodations. A student's ability to learn is likely influenced by his/her home life, individual characteristics, educational background, community, and language (Butler & Stevens, 1997). ELL status, for example, has accounted for a significant proportion of the variance in test scores in previous research (Abedi, Leon, & Mirocha, 2000; Albus et al, 2005).

Another variable which was included in this study, socioeconomic status (SES), has been found to be related to the test performance of ELL students. Abedi and Lord (2001) found that ELL students of low SES had lower scores than ELL students of high SES. Furthermore, when looking at only students who were eligible for free or reduced-price lunch, students were further broken down into income groups and students in the lowest income group were found to have the lowest scores (Hakuta, Goto, & Witt, 2000).

As discussed in the first chapter, SES and ELL status are also related (Abedi, 2002; Abedi, Leon, & Mirocha, 2000; Abedi and Lord, 2001; Batalova, Fix, & Murray, 2007; Collier, 1987). As the data in the tables from Chapter Two indicate, with a large

proportion of students being eligible for free and reduced-price lunches, SES is a highly relevant variable for schools.

Because the dataset used for this project included information on whether students were eligible for free or reduced-price lunches, and did not include data on family income, free or reduced-price lunch status was used as a proxy for SES. There is some variation in how schools decide which students will be given free or reduced-price lunches, making this an imperfect indicator of SES, but this method has been used often and is regarded as reasonable (Abedi and Lord, 2001; Hakuta, Goto, & Witt, 2000).

Home language was another variable found to be important in previous research and included in this study. As discussed in the previous chapter, students whose home language is not English tend to score lower on tests (Abedi, Lord et al, 2001; Abedi, Lord, & Plummer, 1997) and have difficulties in school beyond their test scores, such as repeating a grade or dropping out (NCES, 2004). These problems are even more common for students who speak Spanish in the home than for students who speak other languages in the home (Myers & Milne, 1988; NCES, 2004).

The number of times a student has changed schools is another predictor of math and reading achievement (Abedi, Lord, & Hofstetter, 1998). Students who change schools tend to perform more poorly in school (Mitchell, Destino, & Karam, 1997). Therefore, migrants, who move to new schools frequently, could have even lower scores than other students when other differences are accounted for. Although migrant status and number of schools attended will not give identical results, migrant status was the best available information.

This project used gender as a variable because some research has shown that females get reclassified more quickly than males (Grissom, 2004), which could mean that males and females are different enough to have different rates of efficacy for different accommodations. Other research has not found the efficacy of accommodations to differ for males and females (Albus et al, 2005; Kopriva et al, 2007), but these two studies were

dissimilar to the present study because they manipulated which accommodations students received using random assignment. Furthermore, the Albus et al (2005) study was restricted to Hmong students.

The current study used ethnicity as a variable because it has been shown to be related to school achievement (NCES, 2004), and because it is one of the ways in which states differ considerably from each other, as shown in the previous chapter. Although most previous studies recorded the ethnicities of their participants, they did not treat ethnicity as a variable. Abedi (2008) and Abedi, Leon, and Mirocha (2000) both noted that ELL status could be predicted by ethnicity, but the current study took this a step further and included ethnicity as a descriptor of students receiving testing accommodations.

Abedi, Lord, and Plummer (1997) considered ethnicity as a variable, and found that White, non-Hispanic students scored higher on their math test than Asian students, followed by African American students, then Hispanic students. They did not control for other effects, such as SES or ELL status, however. This study described students using multiple variables in order to provide a broader depiction of the students receiving accommodations.

Another variable of interest is grade level. The research discussed in Chapter Two covered an array of grade levels, but each study only covered one or two grades. As mentioned previously, the gap between ELL and non-ELL students' test performances is larger in the higher grades than the lower grades (Abedi, Leon, & Mirocha, 2000; Butler & Castellon-Wellington, 2000; Hakuta, Goto, & Witt, 2000). In order to provide for a more comprehensive analysis, this project included 3rd through 8th grades.

Analysis

The first research question was “What is the relationship between socioeconomic status, migrant status, ethnicity, gender, grade level, home language, and which

accommodations the ELL students receive?” In order to answer this question, the ELL student group was broken down into the testing accommodation groups they belonged to and percentages of each ethnic group were reported. There were five types of possible accommodations for ELL students in this study: 1) giving extended time; 2) repeating instructions; 3) providing English to native language word-to-word bilingual dictionaries with no definitions; 4) having the ELL teacher administer the test; and 5) reading items aloud. The accommodations were not mutually exclusive.

Next, groups of students receiving each accommodation were described by reporting the percentages for each of the variables. The variables included: grade level (3-8); gender; free or reduced-price lunch eligibility (yes or no); migrant status (yes or no); ethnicity; home language; and type of accommodation. Based on the emphasis found in most states' accommodation documentation that each student should be given an accommodation suited best for that individual, it was expected that the makeup of each accommodation group would differ from the others.

The second research question was “To what extent do Math, Science, and Reading scores vary among ELL students receiving different types of accommodations?” To answer this question, the average percentile ranks of the Reading, Math, and Science tests for each group of students receiving each accommodation were reported. It was expected that the ELL students with the extended time and dictionary accommodations might perform better than the ELL students receiving the other accommodations. Additionally, because Reading scores might be more dependent on English proficiency, it was expected that all ELL students would perform better in Math and Science than in Reading.

Lastly, the third research question was “How does the gap between ELL and non-ELL performance in a norming study compare to low stakes (NAEP) and high stakes (NCLB state assessments) ELL achievement gaps?” Effect sizes were calculated for each content area and grade level comparing the mean for ELL students receiving

accommodations to the overall sample. As a point of comparison, the NAEP and NCLB state assessment results were also used to calculate effect sizes for the gaps between ELLs and the overall sample for each content area and grade level. Because the testing environment in this study is more similar to the current NCLB testing environments in states, it was expected that the effect sizes for the study would be closer to the state assessment effect sizes than to those of the NAEP. The next chapter provides descriptive information about the sample, as well as the results of the analyses described above.

CHAPTER FOUR – RESULTS

This chapter gives the results of the examination of the dataset and the analyses discussed in Chapter Three. The data are first described in terms of the student environment, consisting of school and grade level. The data are next described in terms of the student's individual information, consisting of gender, ethnicity, migrant status, free or reduced-price lunch eligibility, and home language. Lastly, the student percentile ranks from the Reading, Math, and Science tests are described. The results for the entire sample, the entire ELL student group, and ELL students in each accommodation group are presented together for ease of comparison. These results are presented in an effort to answer the following research questions.

- 1) What is the relationship between socioeconomic status, migrant status, ethnicity, gender, grade level, home language, and which accommodations the ELL students receive?
- 2) To what extent do Math, Science, and Reading scores vary among ELL students receiving different types of accommodations?
- 3) How does the gap between ELL and non-ELL performance on a norming study compare to low stakes (NAEP) and high stakes (NCLB state assessments) ELL achievement gaps?

Environment

Some research (Abedi, Lord, & Hofstetter, 1998) has indicated that school attended could be a potent variable in predicting math and reading test scores, and other research indicated it was not an important variable (Abedi, Hofstetter et al, 2001), so it was a factor that warranted investigation for the discussion of testing accommodations. An overall look at how accommodations were distributed among schools is given in Table 6. Of the 195 schools in the study, 87 included ELL students in the study data, with a total of 1,262 ELL students in the study. The 3.8% of ELL students found in this

study was lower than the national percentage of 8.8% listed in Table 3. Of the 87 schools which included ELL students in the study, 45 did not assign any accommodations to ELL students and 42 assigned at least one of their ELL students an accommodation, with a total of 167 ELL students receiving a testing accommodation. Students were not limited to using only one accommodation, but there was only one student who used more than one accommodation, and this student was given three different accommodations.

Table 6. Number of ELL Students Receiving Accommodations per School Having Accommodations

	Number of Schools having accommodations	Number of ELL Students Receiving Accommodations per School				
		Min	Max	Median	Mean	Standard Deviation
Schools	42	1	10	3	4.0	3.0
Extended Time	29	1	8	2	3.0	2.0
Dictionary	6	1	8	1	3.2	3.4
Read-Aloud	2	1	1	1	1	0
Repeated Instructions	19	1	4	1	1.8	1.1
ELL Teacher Administration	9	1	8	1	3	2.9

Although there were some clusters of ELL students receiving the various accommodations at particular schools, the highest being 10 students at one school, it was more often the case that only one or two ELL students at a school received an accommodation. The table also reports how each accommodation was distributed across

schools. The extended time, repeated instructions, and ELL teacher administration accommodations were the most popular, while the dictionary and read-aloud accommodations were only offered by a few schools. It is unclear, based on states' documentation of how accommodation assignments are made, why these accommodations would vary this way in popularity. The extended time and dictionary options were mentioned most often, while small group administration, repeated instructions, and read-aloud were not mentioned as often. Many states encouraged ELL teams to give out accommodations sparingly, so that could account for the small number of ELL students receiving any accommodations.

The distribution of ELL students across schools was also examined. The numbers of ELL students involved in the study from these schools ranged from one to 233, with a median of six, a mean of 14.5, and a standard deviation of 30.0. Similar to the arrangement with testing accommodations, there were some larger groups of ELL students at particular schools, but more often there was a smaller group if there were any at all. More detailed information about the schools which had ELL students can be found in Tables 7-8.

Because of Shepard, Taylor, and Betebenner's (1998) finding that schools with varying proportions of ELL students differ in whether they offer testing accommodations to ELL students, schools were also examined for a pattern between the number of ELL students or the proportion of ELL students and the number of students receiving testing accommodations. These data are given in Tables 7 and 8.

No pattern was evident for the schools with the lowest numbers of ELL students and lowest proportions of ELL students because they did not assign testing accommodations to ELL students in similar numbers. The schools on the low end of the scales, both by number and by proportion of ELL students per school, assigned testing accommodations to students in varying numbers over the whole range (0-10) of the possible total number of ELL students receiving accommodations. The schools with

Table 8. Total ELL Students Receiving Accommodations by Proportion of ELL Students in Schools

Proportion of ELL Students in Study per School	Number of Schools	Number of ELL Students Receiving Accommodations										
		0	1	2	3	4	5	6	7	8	9	10
.00-.01	14	5	5	3	0	0	0	1	0	0	0	0
.01-.05	38	18	1	6	4	2	0	2	3	0	0	2
.05-.10	10	3	2	0	0	1	0	0	1	2	0	1
.10-.20	7	5	1	0	0	1	0	0	0	0	0	0
.20-.30	9	7	0	0	0	0	0	0	0	1	0	1
.30-.40	4	3	1	0	0	0	0	0	0	0	0	0
.40-.50	2	2	0	0	0	0	0	0	0	0	0	0
.50-.61	3	2	1	0	0	0	0	0	0	0	0	0

Next, Table 9 displays the percentages of students from each grade level for the entire sample, the ELL students, the ELL students who received accommodations, and each accommodation group. The discussion of the accommodation groups by grade level, as well as the following discussions regarding gender, ethnicity, home language, free and reduced-price lunch eligibility, and migrant status, are provided in order to answer the research question “What is the relationship between socioeconomic status, migrant status, ethnicity, gender, grade level, home language, and which accommodations the ELL students receive?”

Table 9. Grade Levels for Sample, ELL Students, and Accommodations Groups

	N	Grade Level					
		3	4	5	6	7	8
Entire Sample	33,226	9.0% n=2,979	20.3% n=6,734	20.3% n=6,730	19.9% n=6,618	15.7% n=5,221	14.9% n=4,944
ELL Students	1,262	11.2% n=141	17.6% n=222	19.1% n=241	17.0% n=215	20.4% n=257	14.7% n=186
ELL Students with Accommodations	167	12.0% n=20	18.6% n=31	18.6% n=31	34.7% n=58	12.0% n=20	4.2% n=7
Extended Time	86	11.6% n=10	12.8% n=11	19.8% n=17	44.2% n=38	10.5% n=9	1.2% n=1
Dictionary	19	0% n=0	5.3% n=1	0% n=0	47.4% n=9	15.8% n=3	31.6% n=6
Read-aloud	2	0% n=0	50.0% n=1	0% n=0	50.0% n=1	0% n=0	0% n=0
Repeated Instructions	35	20% n=7	28.6% n=10	22.9% n=8	5.7% n=2	22.9% n=8	0% n=0
ELL Teacher Administration	27	11.1% n=3	29.6% n=8	29.6% n=8	29.6% n=8	0% n=0	0% n=0

For the entire sample, there were not as many 3rd-graders as students in each of the other grades. The ELL group as a whole also had fewer 3rd-graders, but the differences were not as large. This pattern was not reflected in the number of students at each grade level receiving the accommodations, however. As a whole, there were fewer ELL students in grade 8 receiving testing accommodations and more in grade 6.

As noted earlier, extended time was the most commonly offered accommodation. The number of students in grade 8 receiving the extended time condition was very small in comparison with the other grades and in consideration of the distribution of students across grade levels for the entire sample. The distribution of students receiving access to a dictionary across grade levels was unique because, except for one student, only 6th-graders and beyond were given the accommodation. The read-aloud condition results cannot be used to make any inferences due to the sample size. There was a noticeable difference in students receiving repeated instructions, with this accommodation being given fairly equal spread across all grades except 6th and 8th. A chi-square test ($X^2=72.918$, $df=15$, $p<.001$) revealed a significant difference between accommodation groups on grade level. The two people receiving the read aloud accommodation were not included in this statistical analysis, and were not included in the other statistical analyses in this study, because of the small sample size.

It is distinctive that students from 8th grade did not receive accommodations to the degree that students in the other grade levels did. None of them received the repeated instructions accommodation and only one received the extended time accommodation. Neither the 7th- nor 8th-grade ELL students were administered the test by their ELL teachers. Perhaps there is a difference between elementary schools and junior high schools in their ability, willingness, or need to offer accommodations. For example, there may be fewer ELL teachers beyond the elementary level for some schools. There were fewer ELL students in 8th grade, but the difference was not large. However, the English

proficiency levels of the students may have been higher, making it unnecessary to offer an accommodation.

Individual Characteristics

Next, individual characteristics of the sample, ELL students, ELL students who had accommodations, and the specific accommodation groups are discussed. Gender was examined first for the various groups of interest. As can be seen in Table 10, there were roughly as many females as males in the sample, among the ELL students, among the ELL students receiving accommodations, and among most of the accommodation conditions. Not all students provided an indication of gender, so the numbers of students in each group are lower than those existing in the data. There were slightly more males receiving the dictionary and repeated instructions accommodations and slightly more females receiving the extended time and ELL teacher administration accommodations. A chi-square test revealed no significant differences between the genders by accommodation group membership ($\chi^2=3.786$, $df=3$, $p=.285$).

In Table 11 are the percentages of each ethnic group within the sample as a whole, among ELL students, among ELL students receiving accommodations, and in each of the accommodation conditions. When reviewing these percentages, keep in mind that these are unweighted results, whereas the groups were weighted when the final norms were derived. As can be seen in Table 12, the study's percentages in each ethnic group did not entirely match the percentages found for the United States as a whole as displayed in Table 1, but most were reasonably close. The totals are greater than 100% because students were allowed to select more than one ethnicity. There were also students receiving accommodations who did not report an ethnicity.

Table 10. Gender for Sample, ELL Students, and Accommodations Groups

	N	Males	Females
Entire Sample	32,754	50.3% n=16,481	49.7% n=16,273
ELL Students	1,234	51.2% n=632	48.8% n=602
ELL Students with Accommodations	163	46.7% n=78	50.9% n=85
Extended Time	83	44.6% n=37	55.4% n=46
Dictionary	19	63.2% n=12	36.8% n=7
Read-aloud	2	0% n=0	100% n=2
Repeated Instructions	35	57.1% n=20	42.9% n=15
ELL Teacher Administration	26	42.3% n=11	57.7% n=15

Table 11. Ethnicities for Sample, ELL Students, and Accommodations Groups

	N	American Indian/ Alaska Native	Asian	Black/ African American	Hawaiian/ Pacific Islander	Hispanic	White
Entire Sample	33,226	5.4% n=1,790	4.3% n=1,431	18.0% n=5,975	1.0% n=343	22.0% n=7,327	67.6% n=22,444
ELL Students	1,262	3.1% n=39	9.7% n=123	4.7% n=59	1.3% n=17	70.0% n=883	58.9% n=743
ELL Students with Accommodations	167	4.8% n=8	10.8% n=18	22.8% n=38	3.6% n=6	34.1% n=57	49.7% n=83
Extended Time	86	3.5% n=3	5.8% n=5	31.4% n=27	5.8% n=5	25.6% n=22	47.7% n=41
Dictionary	19	10.5% n=2	47.9% n=9	15.8% n=3	5.3% n=1	5.3% n=1	31.6% n=6
Read-aloud	2	0% n=0	0% n=0	50% n=1	0% n=0	0% n=0	50% n=1
Repeated Instructions	35	5.7% n=2	5.7% n=2	11.4% n=4	0% n=0	41.9% n=15	68.6% n=24
ELL Teacher Administration	27	3.7% n=1	7.4% n=2	11.1% n=3	0% n=0	77.8% n=21	48.1% n=13

Table 12. Ethnic Group Comparison between Study Sample and the United States

	American Indian/Alaska Native	Asian/Pacific Islander	Hispanic	Black, non-Hispanic	White, non-Hispanic
Sample	5.4%	5.3%	22.0%	12.3%	43.6%
United States	1.2%	4.9%	21.2%	16.8%	54.3%

The percentage of American Indians/Alaska Natives was lower than 3% for all but one state, New Mexico, which had over 11%, so this proportion was slightly higher than states where previous research has been performed, but not widely different. Because Table 1 combined Asian and Pacific Islander as one group, the percentage to be used to compare the current study's proportion to the studied states and the country as a whole would be 5.3%. Using this percentage, the current study came close to the 4.9% found nationwide. The percentage of Hispanic students for this study was also close to the 21.2% found nationwide.

Because Table 1 referred to only the Black students who are non-Hispanic, the data from the current study were broken down into Black, Hispanic students (0.7% of the sample) and Black, non-Hispanic students (12.3% of the sample). These percentages do not add to the 18.0% found for all Black students in the sample because there were some Black students who did not respond to the question about being Hispanic/Latino. The 12.3% of the sample being Black, non-Hispanic is slightly lower than the national average and most similar to Texas from the states in Table 3.

Similarly, because Table 1 referred to only the White students who are non-Hispanic, the data from the current study were broken down further into White, Hispanic students (14.3% of the sample) and White, non-Hispanic students (43.6% of the sample).

These percentages do not add to the 67.6% found for all White students in the sample because there were some White students who did not respond to the question about being Hispanic/Latino. The 43.6% figure for the number of White, non-Hispanic students is lower than the national average and not similar to any particular state from Table 1. Because there are many ways the data can be broken down into subgroups based on students who selected more than one ethnic group, in addition to whether they were Hispanic or not, further mention of ethnicities will adhere to the overarching ethnic groups.

The proportions of the ethnic groups receiving the extended time accommodation were not very different from the entire sample. The percentage of Black students receiving this accommodation was slightly higher, and the percentage of White students was lower. When compared to the ethnicities in the ELL student group, the extended time group had a much higher proportion of Black students, a much lower proportion of Hispanic students, and a somewhat lower proportion of White students. The proportions were most similar to the overall group of ELL students who received an accommodation, which is to be expected given that the extended time accommodation group had the highest number of ELL students of all the accommodation groups.

There were more Asian students receiving access to a dictionary than would be consistent with their percentage in the sample, ELL group, or the overall group of ELL students who received an accommodation. As with the extended time group, the dictionary group also had fewer White students when compared against the sample and the ELL group, but the proportion was similar to the overall group of ELL students who received an accommodation. The proportion of Hispanic students was much lower than the entire sample, the ELL student group, and the overall group of ELL students who received an accommodation. Although the proportion of Black students in the dictionary group was similar to the proportion found in the entire sample and the overall group of

ELL students who received an accommodation, there was a higher proportion of Black students in this group than the group of ELL students.

The repeated instructions group had slightly more Hispanic students than would be consistent with the sample, but the percentages of the other ethnicities were in keeping with the sample. There was a higher proportion of Black students and a lower proportion of Hispanic students in the repeated instructions group when compared to the ELL group. There were higher proportions of White and Hispanic students, and lower proportions of Asian and Black students, when compared to the whole group of ELL students who received accommodations.

The groups which had the test administered by an ELL teacher had a larger percentage of Hispanic students, along with a slightly smaller percentage of White students, when compared to the total sample. The proportions of Black and Hispanic students were slightly higher for this group than in the ELL group as a whole, and the proportion of White students was slightly lower than the ELL group. There was a lower proportion of Black students, and a much higher proportion of Hispanic students in this group than in the whole group of ELL students receiving accommodations. This accommodation group looked most similar in ethnicity to the ELL group as a whole.

Overall, there were a small number of American Indian ELL students receiving testing accommodations, and there were similar numbers receiving each accommodation. Of the Asian ELL students receiving accommodations, about half were allowed to use a dictionary. Of the Black ELL students receiving accommodations, the majority were given extended time. Of the few Hawaiian ELL students receiving accommodations, all but one received extended time. Of the Hispanic ELL students receiving accommodations, most received extended time or had the test administered by an ELL teacher, and a substantial number received the repeated instructions accommodation. Lastly, about half of the White ELL students receiving accommodations were given extended time, while about 30% were given repeated instructions.

In order to take into account any dependence created by having students with more than one ethnicity in multiple categories, the same data were reviewed by a second method. The students with only one ethnicity selected were assigned to the corresponding ethnic categories, but those who selected more than one ethnicity were assigned to a mixed race category. In addition, the mixed races were reviewed and found to consist of nearly every possible mixture of the six categories. The largest group was the White-Hispanic group, with 4,590 students. The second largest, White-American Indian, was much smaller at 555 students. The percentages for each of the ethnicity categories, including White-Hispanic, for the sample, ELL group, group of ELLs with accommodations, and each accommodation group are given in Table 13. A chi-square test revealed significant differences between the accommodation groups by ethnicity both when using the mixed race category ($\chi^2=67.639$, $df=18$, $p<.001$) and when using the White, Non-Hispanic category with no overlapping of the mixed race category ($\chi^2=81.176$, $df=21$, $p<.001$).

Although the percentages for the American Indian, Asian, Black, and Hawaiian groups changed slightly, the most notable changes occurred in the Hispanic and White groups. This is unsurprising because of the large number of White-Hispanics in the group. It is also notable that one-fifth of the students in the sample were mixed race. Another notable fact is that, although 68% of the mixed race students in the sample were White-Hispanics, and half of the ELL students were mixed race, only 3% of the ELL students were White-Hispanics. Over half of the ELL students receiving accommodations who were mixed race belonged to the White-Hispanic group, though. All of the mixed race students receiving the ELL teacher administration accommodation, most of the mixed race students receiving the repeated instructions accommodation, and half of the mixed race students receiving the extended time accommodation were White-Hispanics. The largest difference is with the dictionary accommodation, which is not

surprising given the large proportion of Asian students in that group. Ethnicity is a complex issue which warrants further attention in test accommodation research.

Table 13. Ethnicities for Sample, ELL Students, and Accommodations Groups with No Overlap

	N	American Indian/ Alaska Native	Asian	Black/ African American	Hawaiian/ Pacific Islander	Hispanic	White	Mixed Race	White Hispanic
Entire Sample	33,226	2.0% n=662	3.2% n=1,077	15.6% n=5,192	0.4% n=127	6.1% n=2,025	49.6% n=16,479	20.2% n=6,711	13.8% n=4,590
ELL Students	1,262	1.1% n=14	8.9% n=112	3.9% n=49	1.0% n=12	22.2% n=280	12.0% n=152	49.0% n=618	2.9% n=36
ELL Students with Accommodations	167	1.2% n=2	10.2% n=17	19.8% n=33	1.2% n=2	9.6% n=16	29.9% n=50	34.7% n=58	20.4% n=34
Extended Time	86	0% n=0	4.7% n=4	24.4% n=21	2.3% n=2	8.1% n=7	30.2% n=26	23.3% n=20	11.6% n=10

Table 13. Continued

	N	American Indian/ Alaska Native	Asian	Black/ African American	Hawaiian/ Pacific Islander	Hispanic	White	Mixed Race	White Hispanic
Dictionary	19	5.3% n=1	47.4% n=9	15.8% n=3	0% n=0	0% n=0	15.8% n=3	15.8% n=3	5.3% n=1
Read-aloud	2	0% n=0	0% n=0	50% n=1	0% n=0	0% n=0	50% n=1	0% n=0	0% n=0
Repeated Instructions	35	0% n=0	5.7% n=2	11.4% n=4	0% n=0	2.9% n=1	34.3% n=12	37.1% n=13	31.4% n=11
ELL Teacher Administration	27	3.7% n=1	7.4% n=2	11.1% n=3	0% n=0	29.6% n=8	0% n=0	44.4% n=12	44.4% n=12

Moving on to further description of the accommodations groups using another of the chosen variables, the percentages of students who selected English, Spanish, or Other for their home language are listed in Table 14. The percentages do not add up to 100% because there were some students who left all three options blank, and there were also others who selected two or three of the options. A chi-square test revealed a significant difference between accommodation groups on home language ($X^2=52.272$, $df= 6$, $p<.001$).

As expected, there were proportionally fewer students from the group speaking English at home in the ELL group and in all of the accommodation groups than in the entire sample. Surprisingly, however, the group of ELL students receiving the accommodations had different proportions of home languages from the ELL group as a whole. Both groups had higher proportions of students who spoke Spanish or another language besides English at home when compared to the overall sample, but the group receiving accommodations had a higher proportion of students speaking English at home and a lower proportion of students speaking Spanish at home when compared to the ELL group as a whole. There were similar proportions speaking another language at home besides English or Spanish for both the group of ELL students and the group of ELL students receiving accommodations.

There were higher proportions of students speaking Spanish at home in the accommodation groups receiving the extended time, repeated instructions, and ELL teacher administration accommodations than in the entire sample. There were higher proportions of students speaking other languages at home in the accommodations groups receiving the extended time, dictionary, repeated instructions, and ELL teacher administration accommodations.

Table 14. Home Languages for Sample, ELL Students, and Accommodations Groups

	N	English	Spanish	Other
Entire Sample	33,226	77.0% n=25,589	10.7% n=3,570	3.1% n=1,041
ELL Students	1,262	17.5% n=221	47.2% n=596	20.1% n=254
ELL Students with Accommodations	167	46.7% n=78	29.9% n=50	22.2% n=37
Extended Time	86	61.6% n=53	22.1% n=19	12.8% n=11
Dictionary	19	31.6% n=6	5.3% n=1	68.4% n=13
Read-aloud	2	50% n=1	0% n=0	50% n=1
Repeated Instructions	35	42.9% n=15	40% n=14	17.1% n=6
ELL Teacher Administration	27	11.1% n=3	66.7% n=18	22.2% n=6

There were higher proportions of students in the extended time, dictionary, and repeated instructions accommodations speaking English at home than in the overall ELL student group, similar to the whole group of ELL students receiving accommodations. In contrast, there was a smaller proportion of students speaking English at home in the ELL teacher administration group than in the overall ELL student group. The proportions of students speaking Spanish at home have the reverse pattern, with the extended time, dictionary, and repeated instructions groups having smaller proportions than the ELL group as a whole and the ELL teacher administration group having a higher proportion than the ELL group. The group receiving the extended time accommodation had a similar proportion of students speaking Spanish at home when compared to the whole group of ELL students receiving accommodations. The dictionary accommodation group had a very small proportion of students who spoke Spanish at home. The proportion of students speaking Spanish at home in the repeated instructions group was most similar to the ELL group.

The pattern for having another language spoken in the home is different. The extended time group had a slightly lower proportion than the ELL group or the ELL group receiving accommodations. The dictionary group had a much larger proportion than the two ELL groups. The repeated instructions and ELL teacher administration groups had very similar proportions to the two ELL groups as a whole.

Overall, of the ELL students receiving accommodations who spoke English at home, most received extended time. Of the ELL students receiving accommodations who spoke Spanish at home, the majority received extended time or had the test administered by an ELL teacher. Of the ELL students receiving accommodations who spoke another language at home, most were allowed to use a dictionary or given extended time.

The percentage of migrants was low for the sample as a whole. There was a slightly higher proportion of migrants in the ELL student group in comparison, but there were only two migrant ELL students receiving testing accommodations.

Table 15 displays the percentages of students eligible for free or reduced-price lunch in the sample, the ELL group, and the accommodations groups. The sample had a smaller percentage of students eligible for free or reduced-price lunches than has been found nationally, as was displayed in Table 2. There were slightly more free or reduced-price eligible students in the overall group of ELL students receiving accommodations and the extended time group, some more yet in the group of ELL students as a whole and the repeated instructions group, and quite a bit more in the ELL teacher administration accommodation group, than there were proportionally in the entire sample. There were no students eligible for the free or reduced-price lunch program receiving the dictionary or read-aloud accommodations. A chi-square test ($X^2=9.560$, $df=3$, $p<.05$) revealed that there was a significant difference between the groups on free or reduced-price lunch eligibility.

Table 15. Free or Reduced-Price Lunch Eligibility for Sample, ELL Students, and Accommodations Groups

	N	FRPL Eligible
Entire Sample	33,226	18.1% n=6,027
ELL Students	1,262	31.2% n=394
ELL Students with Accommodations	167	22.2% n=37
Extended Time	86	22.1% n=19
Dictionary	19	0% n=0
Read-aloud	2	0% n=0
Repeated Instructions	35	25.7% n=9
ELL Teacher Administration	27	40.7% n=11

Test Scores

Given the environmental and individual characteristics for the sample, it was important to investigate the test performances of the groups of interest in this study. The following analyses are provided in order to answer the research question “To what extent do Math, Science, and Reading scores vary among ELL students receiving different types of accommodations?” The national percentile ranks, which come from within-grade analyses, for the Reading test from the sample, the ELL group, the ELL students receiving accommodations, and the different accommodation groups are detailed in Table 16. The number of students in each group may be lower than the number of students for that group in the dataset because not all students had scores for all tests.

The ELL students had a range similar to the sample as a whole, but had a much lower mean percentile rank. The ELL students receiving accommodations as a whole also had a lower mean percentile rank than the sample, but it was slightly higher than the ELL group as a whole. The students receiving the extended time accommodation did not do as well on average compared to the entire sample, but they performed somewhat better than the ELL student group as a whole and slightly better than the group of ELL students receiving accommodations. Those receiving the dictionary accommodation had no students attaining percentile ranks above 74, although their mean percentile rank was the highest out of all of the accommodation groups and much higher than the group of ELL students receiving accommodations, almost equal to that of the entire sample. There was only one student receiving the read-aloud accommodation who had a score for the Reading test, and that student’s percentile rank was very low. Not only was the top percentile rank for the students in the repeated instructions group lower than the entire sample, but the mean was lower as well. However, the mean percentile rank for the repeated instructions group was slightly higher than the mean for the ELL group, similar to the ELL students receiving accommodations. The students who were administered the

test by an ELL teacher had both a very low maximum percentile rank and a much lower average percentile rank than the sample and somewhat lower than the ELL group.

Table 16. Reading National Percentile Ranks for Sample, ELL Students, and Accommodations Groups

	N	Minimum	Maximum	Median	Mean	Standard Deviation
Entire Sample	32,543	1	99	46	48.7	28.9
ELL Students	1,240	1	98	18	25.5	22.4
ELL Students with Accommodations	164	1	97	23.5	31.9	26.4
Extended Time	86	1	97	26	34.7	27.9
Dictionary	19	2	74	46	42.8	24.0
Read-aloud	1	9	9	9	9	N/A
Repeated Instructions	33	2	84	21	30.5	26.9
ELL Teacher Administration	27	1	46	13	18.3	14.7

The percentile ranks for the Math test for the entire sample, the ELL students, and each accommodation group can be found in Table 17. As with the Reading test, the ELL group had a similar range, but lower mean percentile rank than the sample as a whole. Once again, the group of ELL students receiving accommodations had a mean percentile rank lower than the sample, but higher than the ELL student group as a whole. Also similar to the performance on the Reading test, the students in the extended time

condition had a lower mean on the Math test than the entire sample, but it was higher than the mean for the ELL group, similar to the group of ELL students receiving accommodations. The average percentile rank for the students in the dictionary condition was not only higher than the mean for the ELL group receiving accommodations, but also higher than the average for the sample as a whole. The students in the read-aloud condition had a low average percentile rank, below the mean for the ELL group as a whole. The maximums for the repeated instructions and ELL teacher administration groups were slightly lower than the sample's maximum percentile rank. The mean for the repeated instructions group was higher than the mean for the ELL student group as a whole, but not as high as the mean for the sample, similar to the group of ELL students receiving accommodations. The mean for the ELL teacher administration accommodation group was also lower than the sample mean, and slightly lower than the mean for the ELL group.

The percentile rank information for the last test used for this study, Science, can be found in Table 18. As with the previous two tests, the ELL student group had a similar range as the entire sample, but a lower mean percentile rank, and the group of ELL students receiving accommodations had a mean percentile rank between the sample and the ELL student group as a whole. The students in the extended time condition had a mean percentile rank higher than the ELL group, but lower than the sample, similar to the group of ELL students receiving accommodations, just like with the Reading and Math tests. The students who had dictionaries available to them during the test had a higher mean percentile rank than the whole group of ELL students receiving accommodations and the entire sample, just like with the Math test. The student receiving the read-aloud accommodation had a low percentile rank. The repeated instructions accommodation group and the ELL teacher administration accommodation group both had lower means compared to the sample, but similar to the ELL group. The

ELL teacher administration group also had a lower maximum percentile rank than the sample.

Table 17. Math National Percentile Ranks for Sample, ELL Students, and Accommodations Groups

	N	Minimum	Maximum	Median	Mean	Standard Deviation
Entire Sample	32,461	1	99	49	48.9	28.3
ELL Students	1,228	1	99	22.5	29.9	24.5
ELL Students with Accommodations	165	1	99	29	36.4	27.9
Extended Time	85	1	99	24	35.3	29.5
Dictionary	19	2	98	52	55.3	26.0
Read-aloud	2	4	20	12	12.0	11.3
Repeated Instructions	34	1	87	38	38.6	25.5
ELL Teacher Administration	27	1	82	22	26.1	19.5

It was noticeable that the average percentile ranks for each group from Tables 16-18 were lowest for the Reading test. The sample as a whole did not have a difference because the percentile ranks were based on the entire sample. The results for all three tests are shown together for ease of viewing in Table 19. To see if these differences might be statistically significant for the ELL group as a whole and for the ELL students with accommodations, repeated measures tests were run to look for significant differences in performance on the three tests. The results for the ELL group as a whole

($F=38.993$, $df=2$, 2414, $p<.001$) and the group of ELL students with accommodations ($F=10.425$, $df=2$, 320, $p<.001$) suggest that there is a difference in an ELL student's percentile rank depending on the particular test.

Table 18. Science National Percentile Ranks for Sample, ELL Students, and Accommodations Groups

	N	Minimum	Maximum	Median	Mean	Standard Deviation
Entire Sample	32,576	1	99	49	49.0	28.4
ELL Students	1,230	1	99	23	29.0	23.1
ELL Students with Accommodations	162	1	99	33	37.5	26.9
Extended Time	85	1	99	32	38.6	29.3
Dictionary	18	5	94	58	53.1	27.2
Read-aloud	1	16	16	16	16	N/A
Repeated Instructions	33	1	95	29	33.5	23.1
ELL Teacher Administration	27	1	69	26	30.9	18.8

Pairwise comparisons showed that the ELL students who had accommodations had a performance level on the Reading test that was significantly lower than the level of performance on the Math and Science tests ($p<.010$ and $p<.001$, respectfully), but that the Math and Science tests did not have significantly different results ($p=1.00$). The same pattern was found for the group of ELL students as a whole ($p<.001$ for the differences

between Reading and Math or Science scores, $p=.303$ for the difference between Math and Science scores).

Table 19. Reading, Math, and Science National Percentile Ranks for ELL Students and ELL Students with Testing Accommodations

		Median	Mean	Standard Deviation
ELL Students	Reading	18	25.5	22.4
	Math	22.5	29.9	24.5
	Science	23	29.0	23.1
ELL Students with Accommodations	Reading	23.5	31.9	26.4
	Math	29	36.4	27.9
	Science	33	37.5	26.9

In order to further investigate the variables used in this study, the Reading, Math, and Science percentile ranks were examined in terms of home language and SES. The percentile ranks for ELL students and ELL students receiving testing accommodations who speak English, Spanish, and other languages at home are listed in Table 20 for the Reading test, Table 21 for the Math test, and Table 22 for the Science test.

The average percentile ranks for all three tests are similar, with the English and Other home language groups having results similar to each other and the group of students who speak Spanish at home having lower mean percentile ranks. Both the English and Other groups had mean percentile ranks slightly higher than the ELL group,

but lower than the sample as a whole on the Reading and Science tests. Their mean percentile ranks were much closer to the sample mean than the ELL group mean on the Science test, however. On the Math test, the group that spoke English at home had a mean somewhat lower than the sample, but the group that spoke another language besides English or Spanish actually had a mean percentile rank slightly higher than that of the sample.

Table 20. Reading National Percentile Ranks by Home Language for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	216	1	98	32.4	26.8
	Spanish	587	1	96	22.8	18.7
	Other	252	1	97	29.9	25.6
ELL Students with Accommodations	English	78	1	97	37.4	31.0
	Spanish	50	1	55	20.9	13.3
	Other	35	1	97	37.4	25.9

After removing the 231 ELL students, 12 of whom were receiving accommodations, who listed no languages or more than two languages as their home language, ANOVAs comparing each home language group to the others were performed to determine whether the observed differences were statistically significant. The new

means and standard deviations for each language group are given in Tables 23-25 for each subject area test.

Table 21. Math Percentile Ranks by Home Language for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	218	1	99	34.3	28.0
	Spanish	575	1	95	25.9	20.6
	Other	253	1	99	38.5	28.6
ELL Students with Accommodations	English	76	1	99	40.4	31.1
	Spanish	50	1	57	23.2	14.3
	Other	37	1	99	49.8	26.8

For the ELL group as a whole, all three tests showed significant differences between the three home language groups. For Reading, $F=16.615$, $df=2$, 1012, $p<.001$. For Math, $F=23.723$, $df=2$, 1003, $p<.001$. For Science, $F=17.585$, $df=2$, 1004, $p<.001$. The group of ELL students with accommodations also showed significant differences between the three home language groups. For Reading, $F=5.774$, $df=2$, 150, $p<.01$. For Math, $F=9.627$, $df=2$, 150, $p<.001$. For Science, $F=5.162$, $df=2$, 148, $p<.01$. Pairwise comparisons were run to find which home language groups were significantly different for each subject area. For the Reading performance of all ELL students, those who spoke English at home and those who spoke another language at home had significantly higher

percentile ranks than those who spoke Spanish at home ($p < .001$ and $p < .01$, respectively). Those who spoke English did not have significantly different percentile ranks than those who spoke another language at home ($p = .285$). The pattern was the same for Math ($p < .001$ for the comparisons between Spanish speakers and the two other groups, $p = .242$ for the comparison between English speakers and Other language speakers) and for Science ($p < .001$ for the comparisons between Spanish speakers and the two other groups, $p = .846$ for the comparison between English speakers and Other language speakers).

Table 22. Science National Percentile Ranks by Home Language for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	219	1	99	33.8	26.6
	Spanish	581	1	97	26.0	19.6
	Other	247	1	99	35.6	26.3
ELL Students with Accommodations	English	77	1	99	42.5	29.8
	Spanish	50	1	65	26.7	16.0
	Other	34	1	99	44.5	28.1

Table 23. Reading National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	197	1	98	32.5	27.0
	Spanish	575	1	96	22.9	18.8
	Other	243	1	97	29.3	25.4
ELL Students with Accommodations	English	73	1	93	36.2	31.0
	Spanish	50	1	55	20.9	13.3
	Other	30	1	91	34.3	24.7

Table 24. Math National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	199	1	99	34.4	28.2
	Spanish	563	1	95	26.0	20.6
	Other	244	1	99	38.1	28.8
ELL Students with Accommodations	English	71	1	99	38.9	31.6
	Spanish	50	1	57	23.2	14.3
	Other	32	9	99	48.1	28.3

Table 25. Science National Percentile Ranks by Home Language, with No Overlap, for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
ELL Students	English	200	1	99	34.1	26.6
	Spanish	569	1	97	26.1	19.7
	Other	238	1	99	35.3	26.1
ELL Students with Accommodations	English	72	1	99	41.0	30.0
	Spanish	50	1	65	26.7	16.0
	Other	29	2	99	41.2	28.3

Results were similar for the ELL students who received accommodations. For Reading, those who spoke English at home had significantly higher percentile ranks than those who spoke Spanish at home ($p < .01$). The difference between those who spoke another language at home and those who spoke Spanish at home was not statistically significant, however ($p = .060$). As expected, those who spoke English did not have significantly different percentile ranks than those who spoke another language at home, ($p = .937$). The pattern for the ELL students who received accommodations was the same as with the whole ELL group, however, for Math ($p < .01$ for the comparison between Spanish speakers and the English group, $p < .001$ for the comparison between Spanish speakers and the Other group, and $p = .240$ for the comparison between English speakers and Other language speakers) and for Science ($p < .01$ for the comparisons between Spanish speakers and the English group, $p < .05$ for the comparison between Spanish

speakers and the Other group, and $p=1.00$ for the comparison between English speakers and Other language speakers).

Next, the percentile ranks were parsed for students eligible for the free or reduced-price lunch (FRPL) program. These results are given in Table 26. All three tests had similar results for the FRPL students, with mean percentile ranks lower than the ELL group as a whole. In order to analyze whether the differences between FRPL and non-FRPL students were significant, t-tests were performed for each content area test. The means and standard deviations for the non-eligible students are also presented in Table 26 for comparison. For the overall group of ELL students, significant differences were found between FRPL-eligible and non-eligible students. For Reading, $t=4.800$, $df=1238$, $p<.001$. For Math, $t=6.374$, $df=1226$, $p<.001$. For Science, $t=6.542$, $df=1228$, $p<.001$. Similarly, significant results were found for the group of ELL students with accommodations. For Reading, $t=3.151$, $df=162$, $p<.01$. For Math, $t=3.997$, $df=163$, $p<.001$. For Science, $t=3.631$, $df=160$, $p<.001$. These results show that the FRPL-eligible ELL students scored significantly lower on all three tests when compared to non-eligible ELL students whether they had an accommodation or not.

High vs Low Stakes

When students are being assessed, the situation surrounding the assessment can be different enough that it will affect motivation of the student, as well as whether an accommodation is assigned. In a low stakes test, such as the National Assessment of Educational Progress (NAEP), teachers, parents, and test coordinators, for example, may not be as concerned with assigning the appropriate accommodation. They might be interested in seeing the results without any accommodation. On a high stakes test, however, such as a state assessment for NCLB accountability, state, district, and school personnel may have higher motivation to insure that students do well. They may be spending more time motivating and preparing students to maximize performance. ELL

students might be more apt to receive accommodations on a high stakes test than a low stakes test.

Table 26. Reading, Math, and Science National Percentile Ranks for FRPL Eligible and Non-Eligible Students for ELL Students and ELL Students with Testing Accommodations

		N	Minimum	Maximum	Mean	Standard Deviation
FRPL Eligible						
ELL Students	Reading	384	1	97	20.9	18.1
	Math	380	1	99	23.3	20.2
	Science	384	1	99	22.7	17.9
ELL Students with Accommodations	Reading	37	1	55	20.2	15.8
	Math	37	1	72	20.9	16.7
	Science	37	1	62	23.9	13.0
FRPL Non-Eligible						
ELL Students	Reading	856	1	98	27.5	23.8
	Math	848	1	99	32.8	25.6
	Science	846	1	99	31.8	24.5
ELL Students with Accommodations	Reading	128	1	97	35.3	27.9
	Math	128	1	99	40.9	28.9
	Science	125	1	99	41.6	28.7

In order to compare the gap between scores for ELL and non-ELL students on low and high stakes tests, effect sizes for Reading, Math, and Science were calculated for grades 3-8 using the most recent public data available on state assessment proficiency results for 28 states. The other states did not have their full NCLB report card information available online for ELL student results for all grades. For Science, most states did not give an assessment at all grade levels, so there are fewer data points (ranging from 4 states to 19 states) at each grade level than there are for Reading or Math. The results were all expressed in terms of the percentage of students in each group who had attained proficiency or beyond on that test. The minimum and maximum percentages of students attaining proficiency or beyond, as well as the minimum and maximum sample sizes, are shown in Table 27 for the overall samples and in Table 28 for the ELL subgroups.

Because all of these states reported the overall results and the ELL group results, but not all of them reported the non-ELL results separately, the effect sizes, shown in Table 29, were calculated comparing the overall results to the ELL group results. This means that the ELL group's results are affecting the overall results, and a separate display of the non-ELL results would cause slightly different gap data, most likely larger. Effect sizes were computed using the formula: $d = (M_o - M_e)/s$, where s is the pooled standard deviation, M_o is the mean for the overall results, and M_e is the mean for the ELL student results. Out of all of the effect sizes, only one, Grade 4 Science, had a confidence interval that included zero. The rest had gaps that were significantly higher than zero, meaning overall performance was significantly better than ELL performance.

Table 27. Minimum and Maximum Percentages of Proficiency and Above on State Assessments for Overall Samples

		Reading		Math		Science	
		Min	Max	Min	Max	Min	Max
Grade 3	Percent Proficient	43	94	36	88	41	81
	Sample Size	6,006	462,044	5,945	461,751	20,236	128,326
Grade 4	Percent Proficient	48	89	40	90	38	96
	Sample Size	5,977	467,678	5,939	467,400	6,193	197,303
Grade 5	Percent Proficient	46	94	40	92	15	91
	Sample Size	6,120	463,136	6,064	462,884	20,335	133,778
Grade 6	Percent Proficient	48	94	37	83	41	77
	Sample Size	6,165	464,321	6,093	463,993	9,281	125,588
Grade 7	Percent Proficient	48	91	37	89	40	82
	Sample Size	6,325	467,687	6,332	202,189	55,543	123,953
Grade 8	Percent Proficient	45	97	29	86	16	90
	Sample Size	6,465	201,371	6,465	474,681	6,500	175,068

Table 28. Minimum and Maximum Percentages of Proficiency and Above on State Assessments for ELL Samples

		Reading		Math		Science	
		Min	Max	Min	Max	Min	Max
Grade 3	Percent Proficient	22	93	22	80	38	73
	Sample Size	167	158,340	167	158,276	167	12,902
Grade 4	Percent Proficient	18	81	18	81	8	92
	Sample Size	71	134,181	71	467,400	71	17,986
Grade 5	Percent Proficient	13	90	16	88	2	74
	Sample Size	40	110,153	40	110,121	40	9,510
Grade 6	Percent Proficient	6	83	10	72	31	62
	Sample Size	29	91,734	30	91,684	30	5,514
Grade 7	Percent Proficient	3	69	10	79	18	63
	Sample Size	34	87,189	34	467,308	32	4,250
Grade 8	Percent Proficient	2	83	11	68	3	70
	Sample Size	27	475,746	27	81,401	27	12,312

NAEP state score results (Institute of Educational Sciences, 2012) for Reading, Math, and Science were also used to calculate effect sizes for grades 4 and 8. There were some states which had no Science results for grade 4, but all had results for grade 8 Science and both grades 4 and 8 in Reading and Math. Within each grade level and content area combination, some states did not have enough ELL students tested to calculate a meaningful average score for the state. The minimum and maximum means

and sample sizes for NAEP results are shown in Table 30 for the overall samples and in Table 31 for the ELL subgroups. For comparison purposes, the effect sizes, given in Table 32, were derived by contrasting the average scores for ELL students with those for the overall group of students, even though the non-ELL student averages were available. Similar to the NCLB results, the NAEP results showed no effect size confidence intervals containing zero. These results provide evidence for significantly higher performance levels from the overall samples when compared to the ELL subgroups.

Overall, the NAEP effect sizes for the gap between ELL and non-ELL students were similar to the effect sizes for the NCLB results. Their confidence intervals overlap for all content area and grade level combinations, so there is no evidence of the average effect over all state testing results being significantly different from the average effect for the NAEP testing results.

There was a trend for both sets of data showing that the effect sizes increased, meaning the gap between ELLs and non-ELLs became larger, as grade level increased. For the NAEP, because the confidence intervals overlap between grade 4 and grade 8 effect sizes for each content area, however, the differences are not large enough to be significant. Similarly, for the state assessments' average effect sizes, the differences between any two grade levels for any content area are not large enough to be significant because the confidence intervals overlap. If the effect sizes had been based on the difference between the non-ELL students and ELL students, instead of the overall sample being compared to the ELL student group, it can be assumed that the gap would have been larger because ELL scores, which are lower, would not have been included in the averages. Because of the large overlap in the confidence intervals, however, it is possible that the trend of gaps growing larger as grade level increased would still not be significant.

Gap Comparison

In order to answer the third research question, “How does the gap between ELL and non-ELL performance on a norming study compare to low stakes (NAEP) and high stakes (NCLB state assessments) ELL achievement gaps?”, the effect sizes comparing the ELL students to the entire sample, to be consistent with the comparison made for the NAEP and state assessments, were computed and are shown in Table 33. As expected, non-ELL students in the current study performed better than ELL students on all three tests. This table also demonstrates that there may be a trend for the gap to be larger as grade level increases.

The results of this study are similar to the results of the NAEP and state assessment results in that the confidence intervals between grade levels for each content area overlapped. The only exceptions were found between Grade 3 and Grade 7, with the confidence intervals for Grade 7 encompassing larger gaps than Grade 3. This is evidence that there could be a larger gap between ELL and non-ELL students at higher grade levels than lower grade levels.

When the effect sizes for this study are compared to the NCLB and NAEP results, the confidence intervals have overlap in every grade level and content area combination, meaning there is no evidence that this study produced ELL gap results different from the state assessments or NAEP. For the Grade 8 NAEP Science confidence interval, however, the confidence intervals just barely overlap.

Summary of Accommodation Groups

The accommodation groups varied in their composition by the subgroups based on test scores, grade level, ethnicity, home language, and SES. There were at times larger percentages of certain subgroups receiving an accommodation. There were also instances when a smaller subgroup had a higher proportion of students receiving the accommodation than was consistent with their proportion in the sample. There were only

two students in the read-aloud accommodation, both females, one in 4th grade and one in 6th grade, one Black and one White, one who spoke English at home and one who spoke a language other than English or Spanish at home, neither migrants, neither free or reduced-price eligible, and both with low percentile ranks on the tests for which they had scores.

More students received extended time than any other accommodation. The students receiving the extended time accommodation were mostly White, but there were also large numbers of Black and Hispanic students receiving the accommodation. Most spoke English at home, were not migrants, were not predominantly one gender, and were mostly in 6th grade. The smallest number was in 8th grade, and the rest were split up close to evenly among the other grade levels. Only 20% were free or reduced-price lunch eligible. The mean percentile ranks for this group on the three tests used in this study were lower than the means for the sample, but higher than the means for the ELL group and similar to the whole group of ELL students receiving accommodations.

The majority of those receiving the dictionary accommodation was Asian, but there were also a number of White students receiving this accommodation. Most spoke another language besides Spanish or English at home. None of the members of this group was migrant or in the free or reduced-price lunch program. There were similar numbers of males and females. Most were in 6th grade, followed by 8th grade, then 7th grade. There was only one student receiving this accommodation from grades 3-5. Although the mean percentile rank for this accommodation group was slightly lower than the means for the sample on the Reading test, the mean percentile rank was actually higher than the sample on the Math and Science tests. The average percentile rank was higher for the dictionary group than the ELL group on all three tests.

Table 29. Minimum and Maximum Effect Sizes and Average Effect Sizes for the Gap Between ELL and Overall Percentages of Proficiency and Above on State Assessments

		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
Reading	Min/Max Effect Sizes	0.17	2.07	0.08	2.71	0.33	2.93	0.24	3.23	0.40	3.03	0.24	3.73
	Average Effect Size	1.04		1.22		1.43		1.67		1.78		2.26	
	Confidence Interval	0.18	1.90	0.33	2.11	0.44	2.42	0.65	2.70	0.72	2.85	0.64	3.87
Math	Min/Max Effect Sizes	-0.12	1.84	0.04	2.23	0.10	2.46	0.05	2.69	0.09	2.48	0.06	2.26
	Average Effect Size	0.75		0.83		1.01		1.18		1.37		1.25	
	Confidence Interval	0.02	1.48	0.13	1.53	0.21	1.80	0.39	1.97	0.53	2.21	0.57	1.93
Science	Min/Max Effect Sizes	0.12	0.52	0.21	1.98	0.31	1.96	0.27	1.49	0.33	1.78	0.20	2.06
	Average Effect Size	0.45		1.18		1.24		1.05		1.19		1.25	
	Confidence Interval	0.23	0.66	-0.04	2.40	0.47	2.00	0.36	1.74	0.23	2.15	0.30	2.19

Table 30. Minimum and Maximum Mean Scores on NAEP for Overall Samples

		Reading		Math		Science	
		Min	Max	Min	Max	Min	Max
Grade 4	Mean Score	208	237	231	253	136	163
	Sample Size	2,800	10,400	2,800	10,000	2,500	7,600
Grade 8	Mean Score	255	275	273	299	137	162
	Sample Size	2,500	7,700	2,500	7,900	2,600	7,400

Table 31. Minimum and Maximum Mean Scores on NAEP for ELL Samples

		Reading		Math		Science	
		Min	Max	Min	Max	Min	Max
Grade 4	Mean Score	153	207	201	234	98	143
	Sample Size	56	2,976	50	2,883	58	2,204
Grade 8	Mean Score	211	251	227	272	86	136
	Sample Size	37	1,224	37	1,258	37	1,406

Table 32. Minimum and Maximum Effect Sizes and Average Effect Sizes for the Gap Between ELL and Overall Mean Scores on the NAEP

		Grade 4		Grade 8	
Reading	Min/Max Effect Sizes	0.22	1.61	0.28	1.92
	Average Effect Size	0.91		1.08	
	Confidence Interval	0.58	1.24	0.73	1.43
Math	Min/Max Effect Sizes	0.11	1.33	0.40	1.52
	Average Effect Size	0.79		1.03	
	Confidence Interval	0.52	1.06	0.72	1.34
Science	Min/Max Effect Sizes	0.27	1.49	0.62	1.92
	Average Effect Size	0.92		1.20	
	Confidence Interval	0.62	1.21	0.84	1.56

The majority of students receiving the ELL teacher administration accommodation was Hispanic and spoke Spanish at home. There was not a significantly different number of males and females. There were no migrants in this group. There was an equal spread of students among the 4th, 5th, and 6th grades, with a smaller number in 3rd grade. Approximately one-third of this group was eligible for free or reduced-price lunch. The mean percentile ranks for this accommodation group were much lower than those of the whole sample on all three tests. The mean percentile rank for the Reading test was lower than mean for the ELL group, but mean percentile ranks for the Math and Science tests were similar to the ELL group.

When an overall comparison between ELL students and the sample as a whole was made by calculating the effect sizes for each grade level and content area, the results showed that ELL students had lower test scores than the overall sample. In addition, this gap was significantly larger for Grade 7 than Grade 3.

Table 33. Effect Sizes and Confidence Intervals for the Gap between ELL and Overall Average Percentile Ranks of the Study Sample

		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
Reading	Effect Size	0.70		0.76		0.79		0.70		1.00		0.90	
	Confidence Interval	0.53	0.87	0.63	0.90	0.66	0.92	0.56	0.83	0.87	1.13	0.75	1.05
Math	Effect Size	0.51		0.55		0.68		0.71		0.82		0.75	
	Confidence Interval	0.34	0.68	0.42	0.69	0.55	0.81	0.57	0.85	0.69	0.95	0.60	0.90
Science	Effect Size	0.50		0.72		0.73		0.67		0.84		0.70	
	Lower Limit	0.33	0.67	0.58	0.85	0.60	0.86	0.53	0.81	0.71	0.97	0.56	0.85

As these results show, the characteristics of the students in the different accommodation groups varied. This may be evidence that students are being given accommodations based on their unique situations as opposed to receiving a one-size-fits-all solution. Based on the gaps analysis, the data from this study were similar enough to state assessment data and NAEP data to reinforce previously reported trends in academic achievement. The following chapter will discuss the results from these analyses further and their potential explanations and impact.

CHAPTER FIVE – DISCUSSION

It is important for all students to be tested in a way that allows them to show their abilities without irrelevant issues interfering; thereby enabling teachers to more accurately determine students' strengths and weaknesses. It is also important for assessment information to provide insights into future instructional planning for individual students. To accomplish this, students are sometimes given testing accommodations which are intended to alleviate the irrelevant interference. English Language Learners (ELLs), because of their limited English proficiency, are one of the groups commonly given testing accommodations. Previous research has largely focused on random assignment studies which do not take into account the unique characteristics and circumstances of the students. In this study, the goal was to describe in detail the characteristics of the ELL students being given different types of accommodations in order to investigate whether there were some characteristics which might be important for future research on testing accommodations to consider.

To that end, data from a national sample on a general achievement test battery were used. These data included Reading, Math, and Science test scores, as well as information about grade, gender, ethnicity, migrant status, free or reduced-price lunch eligibility, and home language for students. Each testing accommodation group was compared to the other groups, to the overall group of ELL students, and to the overall sample on the variables. Statistical analyses were used as appropriate to uncover any significant differences between the groups. Additionally, effect sizes were calculated for the gap between the ELL students and the overall group on each test for each grade. These effect sizes were compared to similar effect sizes calculated from NAEP and state NCLB data.

In this study, students were assigned accommodations by their own teachers. The accommodations which were found in this study were (in order of prevalence): being

given extended time, repeating instructions, having an ELL teacher administer the test, using a dictionary, and having the test read aloud. Some of the accommodations mentioned in the literature were not feasible in the collection of data used in this study. These included: computerized testing, linguistic modification, glossaries, and translated forms.

Many of the results found through the analyses were expected based on the research reviewed in Chapter Two. As with most of the previous research, this study had a sample of ELL students that consisted mostly of Hispanic individuals. There was also a larger proportion of Asians found in the ELL group than found in the sample as a whole. Consistent with Myers and Milne (1988), Abedi, Lord, and Plummer (1997), and Abedi, Lord et al (2001), students who spoke Spanish at home had lower percentile ranks on average than students who spoke English or another language for Reading, Math, and Science tests. The students speaking English and Other languages had similar performance levels.

The students who were from low socioeconomic households and were eligible for free or reduced-price lunch did not perform as well as other students. The ELL students did not perform as well as non-ELL students on average. As the states' documentations suggested, the testing accommodations appeared to be assigned based on individual student characteristics, although there was a tendency for students to be given extended time more than other accommodations. Like Albus et al (2005) and Kopriva et al (2007), gender was not found to be a significant factor for assigning accommodations, although Grissom's (2004) finding that females were reclassified as English proficient more quickly suggested that it might be.

Overall, the ELL students receiving accommodations did better than those who did not. The students who were able to use a word-to-word dictionary, similar to Abedi, Lord et al (2001) and Albus et al (2005), had a better performance than other ELL students. Students who had extended time, similar to Abedi, Courtney, and Leon (2003),

Abedi, Hofstetter et al (2001), Garcia (1991), and Hafner (2001), also had a better performance than other ELL students. Exactly opposite to the findings of Abedi, Courtney, and Leon (2003), however, the dictionary group had better performance levels than other accommodations groups.

Meanwhile, the students in the read aloud group had a low performance, which was similar to Castellon-Wellington's (2000) and Shepard, Taylor, and Betebenner's (1998) findings. The students in the ELL teacher administration and repeated instructions groups performed better than the rest of the ELL students, which was not expected because Abedi, Courtney, and Leon (2003) did not show an advantage for that accommodation. It is possible that there are further differences between subgroups that the study was not able to detect, such as proficiency in English or previous schooling in another country. For example, Shepard, Taylor, and Betebenner (1998) found that the difference between scores for students who had extended time and those who did not varied by how proficient in English the students were. Without baseline scores, there is no absolute support for the efficacy of any of these accommodations.

Another unexpected result was the makeup of the dictionary group. Although it was expected that the groups would differ in terms of SES, ethnicity, and home language, it was remarkable how the dictionary group was mostly Asians, most spoke an Other language at home, and none was FRPL eligible. Past research has shown that Asian students generally do better than Hispanic students in school, so perhaps these students were doing well and just needed a little extra help with a word here or there. Perhaps the dictionary accommodation is better suited for students with generally stronger academic skills

Similarly remarkable was the difference in the home language makeup between the ELL group as a whole and the ELL group receiving accommodations. There was a higher percentage of students who spoke English at home receiving accommodations than in the general ELL group. One possible explanation is that parents who are more

proficient at English themselves are more likely to ask for accommodations for their children. If this is true, then this may be a cause of concern for educators.

Finally, an analysis of the gaps between ELL and overall performance levels showed similar results across grades for the NAEP and NCLB test results. The results from this study were also similar to the NAEP and NCLB results, but showed a significantly larger gap for 7th-graders when compared to 3rd graders for all content areas. These results are in agreement with results from Abedi, Leon, and Mirocha (2000), Butler and Castellon-Wellington (2000), and Hakuta, Goto, and Witt (2000) that the gap between ELL and non-ELL students increases as grade level increases and provides support for grade level being an important variable for analysis of testing accommodations.

Summary

This study used several variables – grade level, gender, ethnicity, home language, SES, and migrant status – which other research in testing accommodations has not taken into account. This study was also unique in having testing accommodations assigned by schools instead of randomly by researchers. From these data, it has been found that accommodations may be allocated based on individual characteristics, with each of the accommodations groups having unique characteristics. Ethnicity, grade level, SES, and home language appear to be important variables in regards to deciding which accommodations students will be given. Future testing accommodations research should consider these variables.

Limitations

This study was limited in some obvious and unavoidable ways. The number of ELL students receiving accommodations was relatively small, limiting the nature of the analyses possible. Other studies have attempted to avoid this problem by assigning accommodations to students randomly. Results from this study indicate that such a

strategy risks ignoring potentially important variables and making the results of questionable usage for real world decision-making.

Three variables not used in this study which have been found to be potentially important to predicting success on tests are English proficiency, parental education, and changing schools. English proficiency was found to be a factor in how well ELL students scored, and how much higher their scores were with the use of a testing accommodation (Shepard, Taylor, and Betebenner, 1998). Parental education was found to be a good potential predictor of success in previous studies, having a positive correlation with test scores (Abedi and Lord, 2001; Hakuta, Goto, & Witt, 2000). Research has also found that when students change schools their rate of learning decreases (Mitchell, Destino, Karam, 1997). These variables were not used in this study, however, due to accessibility of the information. Migrant status was used as a proxy for a higher rate of changing schools, but there were few students who were migrants in this sample and it is difficult to separate the effect of changing schools from other effects, such as home language and ELL status. In addition to the variables that previous research has noted may be important, there are also many variables which many states have documented as being factors to consider, such as the student's level of literacy in the native language, the student's difficulty in using the accommodation, the student's willingness to learn to use the accommodation, and the length of time the student has been in the U.S.

Ethnicity was a variable which was not entirely clear cut. The study adhered to the current federal guidelines which dictate that whether students are Hispanic or not is a separate question from their ethnicity, which does not include Hispanic as an option. This structure could be quite confusing for students and cause them to make incorrect selections.

Finally, it was not possible to evaluate the effectiveness of accommodations used by the sample gathered in this study. The relatively small number of ELL students

receiving testing accommodations and a lack of baseline performance precluded this possibility.

Future Research

An area for researchers to focus on in the future is what factors determine which accommodations are given to ELL students. Results in this study and documentation from states indicated that accommodations are assigned depending on different characteristics and circumstances. It would be useful to investigate more precisely how decisions on accommodations are made, such as the importance the decision-making teams place on various factors.

Future studies should try to obtain larger numbers of ELL students, enabling more sophisticated analyses of the factors related to accommodation assignment and effectiveness. More focused sampling of schools with large populations of ELL students would allow hierarchical analyses that could include school level, the grade level, and individual characteristics.

In addition, normally students are not to be given unfamiliar accommodations on the day of assessment. Researchers should be making an effort to familiarize students with the accommodations they are being given before the test is administered.

Final Comments

Students are being given accommodations depending on individual factors, such as English proficiency, native language literacy, time in US schools, age, grade level, and previous schooling in another country. The research on accommodations has heretofore ignored these factors when performing analyses, creating a misalignment between research and practice. Given the current predilection for high-stakes testing and the increasing diversity of populations, it seems likely that more research on testing accommodations will be needed to enable valid test interpretations for ELL students in particular. A consideration of the results of this study might help future researchers to

achieve a more complete understanding of the utility of testing accommodations for ELL students.

REFERENCES

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment*, 8 (3), 231–257.
- Abedi, J. (2004). The No Child Left Behind Act and English language learner: Assessment and accountability issues. *Educational Researcher*, 33 (1), 4–14.
- Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, 27 (3), 17-31.
- Abedi, J., Courtney, M., and Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Tech. Rep. No. 586). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Courtney, M., Mirocha, J., Leon, S. and Goldberg, J. (2005). *Language accommodation for large-scale assessment in science: assessing English language learners* (CSE Tech. Rep. No. 666). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Abedi, J., Hofstetter, C., Baker, E., and Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C. H., and Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74 (1), 1–28.
- Abedi, J., Leon, S., and Mirocha, J. (2000). Examining ELL and non-ELL student performance differences and their relationship to background factors: Continued analyses of extant data. In Abedi, J., Bailey, A., Butler, F.A., Castellon-Wellington, M., Leon, S., and Mirocha, J. *The validity of administering large-scale content assessments to English language learners: an investigation from three perspectives* (CSE Tech. Rep. 663). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Leon, S., and Mirocha, J. (2003). *Impact of students' language background on content-based assessment: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles, CA: Center for the Study of Evaluation/ National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., and Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14 (3), 219–234.
- Abedi, J., Lord, C., and Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Tech. Rep. No. 478). Los Angeles, CA: UCLA Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

- Abedi, J., Lord, C., Kim, C., and Miyoshi, J. (2001). *The effects of accommodations on the assessment of LEP students in NAEP* (CSE Tech. Rep. No. 537). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Lord, C., and Plummer, J. (1997). *Language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No. 429). Los Angeles: University of California, Center for the Study of Evaluation.
- Alabama State Department of Education. (2012). *Alabama's education report card 2010-2011*. Retrieved from <http://www.alsde.edu/general/AlabamaEducationReportCard.pdf>
- Alabama State Department of Education. (2012). *Alabama student assessment program policies and procedures for students of special populations*. Retrieved from <https://docs.alsde.edu/documents/91/alabama%20student%20assessment%20program%20policies%20and%20procedures%20for%20students%20of%20special%20populations.pdf>
- Alaska Department of Education and Early Development. (2011). *Testing accommodations manual for limited English proficient students*. Retrieved from http://eed.alaska.gov/tls/assessment/accommodations/LEP/ELL_Accommodations_Manual_093011.pdf
- Alaska Department of Education and Early Development. (2012). *Consolidated state performance report: Part 1*. Retrieved from http://www.eed.state.ak.us/nclb/pdf/10-11_AK_CSPR_PartI.pdf
- Albus, D., Thurlow, M., Liu, K., and Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research*, 98 (4), 245–253.
- Arizona Department of Education. (2011). *Testing accommodations: Guidelines for school year 2011-2012*. Retrieved from <http://www.azed.gov/standards-development-assessment/files/2011/08/testingaccommodations2011-12.pdf>
- Arizona Department of Education. (2012). *2010-2011 state report card*. Retrieved from <http://www.azed.gov/research-evaluation/files/2012/04/2011stareportcard.pdf>
- Arkansas Department of Education. (n.d.). *LEP accommodations*. Retrieved from <http://arkansased.org/educators/curriculum/lep.html>
- Arkansas Department of Education. (n.d.). *Modifications used in the classroom and accommodations used on state assessments*. Retrieved from www.arkansased.org/testing/pdf/assessment_accommodations.pdf
- Arkansas Department of Education. (2005). *Guidelines for assessment accommodations for students with disabilities*. Retrieved from <http://arksped.k12.ar.us/documents/specialprojects/GuidelinesforAssessmentAccommodations2011.pdf>
- Arkansas Department of Education (2012). *2010-2011 state report card*. Retrieved from <http://normessasweb.uark.edu/schoolperformance/beta/stachievement/index>

- Batalova, J., Fix M., and Murray, J. (2007). *Measures of Change: The Demography and Literacy of Adolescent English Learners—A Report to Carnegie Corporation of New York*. Washington, DC: Migration Policy Institute.
- Bensoussan, M., Sim, D., and Weiss, R. (1984). The effect of dictionary usage of EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language*, 2 (2), 262-276.
- Butler, F. A., and Castellon-Wellington, M. (2000). Students' concurrent performance on tests of English-language proficiency and academic achievement. In Abedi, J., Bailey, A., Butler, F.A., Castellon-Wellington, M., Leon, S., and Mirocha, J. *The validity of administering large-scale content assessments to English language learners: an investigation from three perspectives* (CSE Tech. Rep. 663). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Butler, F. A., and Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: student characteristics and other considerations* (CSE Report No. 448). Los Angeles, CA: University of California, National Center for Research for Evaluation, Standards, and Student Testing.
- California Department of Education. (2011). *State accountability report card 2009-2010*. Retrieved from <http://www.cde.ca.gov/ta/ac/sc/documents/reportcard0910.pdf>
- California Department of Education. (2011). *Testing variations, accommodations, and modifications*. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/sa/documents/matrix2012fnl.doc>
- Castellon-Wellington, M. (2000). *The impact of preference for accommodations: the performance of English language learners on large-scale academic achievement tests* (CSE Technical Report 524). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Collier, V. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, 21 (4), 617-641.
- Colorado Department of Education. (n.d.). *Colorado state report card 2009-2010*. Retrieved from http://www.cde.state.co.us/FedPrograms/dl/danda_nclbstrptcrd_0910.pdf
- Colorado Department of Education. (2011). *Colorado accommodation manual: Selecting and using accommodations for instruction and assessment*. Retrieved from http://www.cde.state.co.us/assessment/documents/coalt/2011_CO_Accom_Manual.pdf
- Connecticut State Department of Education. (2011). *Assessment guidelines section B: Assessing students identified as English language learners*. Retrieved from http://www.csde.state.ct.us/public/cedar/assessment/agl/resources/CT_ELL_Assessment_Guidelines_2011-2012.pdf
- District of Columbia Office of the State Superintendent of Education. (2010). *2009-2010 state report card*. Retrieved from http://osse.dc.gov/sites/default/files/dc/sites/osse/publication/attachments/2009-2010_district_of_columbia_state_report_card.pdf

- District of Columbia Office of the State Superintendent of Education. (2011). *District of Columbia Office of the State Superintendent of Education Testing Accommodation Manual, 2011-2012*. Retrieved from <http://osse.dc.gov/sites/default/files/dc/sites/osse/publication/attachments/2012%20OSSE%20Testing%20Accommodations%20Manual.doc>
- Delaware Department of Education. (n.d.). *Statewide disaggregated DSTP summary*. Retrieved from <http://dstp.doe.k12.de.us/DSTPmart9/SummaryDisaggAllTests.aspx>
- Delaware Department of Education. (2009). *Delaware accommodations manual for ELLs: A guide to selecting, administering, and evaluating the use of accommodations for instruction and assessment of English language learners*. Retrieved from <http://www.doe.k12.de.us/aab/files/2009-10%20ELL%20Accoms%20manual%20with%20fact%20sheets%20and%20teacher%20tools.pdf>
- Fairbairn, S. B., and Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice*, 28 (1), 10-24.
- Florida Department of Education. (2010). *Guide to FCAT and FCAT 2.0 accommodations for students with disabilities*. Retrieved from <http://www.fldoe.org/ese/pdf/fcatteam.pdf>
- Georgia Department of Education. (n.d.). *2010-2011 report card*. Retrieved from <http://www.doe.k12.ga.us/Pages/Report-Card.aspx>
- Georgia Department of Education. (2008). *Accommodations manual: A guide to selecting, administering, and evaluating the use of test administration accommodations for students with disabilities*. Retrieved from <http://archives.gadoe.org/DMGetDocument.aspx/Accommodation%20Manual%20August%202008.pdf?p=6CC6799F8C1371F6D1C69D60E4029F1085A5BB31C1B556F24ADBA65C89A3BD77&Type=D+>
- Grissom, J. B. (2004). Reclassification of English language learners. *Education Policy Analysis Archives*, 12 (36). Retrieved from <http://epaa.asu.edu/epaa/v12n36/>
- Hafner, A. L. (2001). *Evaluating the impact of test accommodations on test scores of LEP students and Non-LEP students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Hakuta, K., Goto, Y., and Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy Rep. No. 2000-1). Santa Barbara, CA: The University of California Linguistic Minority Research Institute.
- Hofstetter, C. H. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education*, 16 (2), 159–188.
- Idaho State Department of Education. (n.d.). *Report card for 2009-2010*. Retrieved from <http://www.sde.idaho.gov/reportcard/>
- Idaho State Department of Education. (2011). *Idaho LEP program manual*. Retrieved from http://www.sde.idaho.gov/site/lep/guidance_docs/Idaho%20LEP%20Program%20Manual%20OCT%202011.doc

- Illinois Department of Education. (2011). *2009-2010 state report card*. Retrieved from http://www.isbe.state.il.us/assessment/pdfs/report_card/rc_11.pdf
- Illinois State Board of Education. (2011). *Assessment accommodations for English language learners guidance for 2011-2012*. Retrieved from http://www.isbe.net/assessment/pdfs/ell_guidance.pdf
- Indiana Department of Education. (2011). *2011-2012 ISTEP+ program manual*. Retrieved from <http://www.doe.in.gov/achievement/assessment>
- Indiana Department of Education. (2012). *ISTEP+ 2011-2012*. Retrieved from <http://compass.doe.in.gov/dashboard/istep.aspx?type=state>
- Iowa Department of Education (2007). *Guidelines for the inclusion of English language learners (ELLs) in K-12 assessment*. Retrieved from http://www.iowa.gov/educate/index.php?option=com_content&view=article&id=683&Itemid=1391#guides
- Iowa Department of Education. (2011). *The state report card for No Child Left Behind 2010-2011*. Retrieved from http://www.educateiowa.gov/index.php?option=com_docman&task=cat_view&gid=670&Itemid=4434
- Iowa Department of Education. (2012). *Iowa guidelines for K-12 ELL participation in districtwide assessments*. Retrieved from http://www.educateiowa.gov/index.php?option=com_docman&task=doc_download&gid=13182&Itemid=4303
- Kansas State Department of Education. (n.d.). *Report card 2010-2011*. Retrieved from <http://online.ksde.org/rcard/summary/state.pdf>
- Kansas State Department of Education. (2011). *Accommodations manual: How to select, administer, and evaluate accommodations for instruction and assessment*. Retrieved from http://www.ksde.org/LinkClick.aspx?fileticket=f_xav-MvMvU%3d&tabid=2372&mid=7983&forcedownload=true
- Kentucky Department of Education. (2009). *Inclusion of special populations in the state-required assessment and accountability programs*. Retrieved from <http://www.education.ky.gov/nr/rdonlyres/c0c41c16-c10d-4022-af71-62c7b8814051/0/inclusiondocument.pdf>
- Kentucky Department of Education. (2012). *Kentucky state report card 2010-2011*. Retrieved from http://www.education.ky.gov/nr/rdonlyres/4c016fa5-ae16-4b2a-93eb-2d0b2d36cb11/0/kyreportcard_2011.pdf
- Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., and Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice* 26 (3), 11-20.
- Liu, K., Anderson, M., Swierzbins, B., and Thurlow, M. (1999). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase I* (Minnesota Rep. No. 20). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

- Louisiana Department of Education. (n.d.). *LEAP and GEE special populations and accommodations*. Retrieved from <http://www.louisianaschools.net/lde/uploads/18146.pdf>
- Louisiana Department of Education. (2010). *Consolidated state performance report part I: For reporting on school year 2009-2010*. Retrieved from <http://www.louisianaschools.net/lde/uploads/18065.pdf>
- Maine Department of Education. (2009). *New England common assessment program accommodations guide*. Retrieved from http://www.maine.gov/education/necap/0910materials/accommodations_guide0909.pdf
- Maine Department of Education. (2010). *Fall 2011 NECAP state results: Grade 3-8 disaggregated results*. Retrieved from <http://www.maine.gov/education/necap/results.html>
- Maryland State Department of Education. (n.d.). *Maryland school assessment results 2010-2011*. Retrieved from <http://www.mdreportcard.org/>
- Maryland State Department of Education. (2008). *Maryland accommodations manual: Selecting, administering, and evaluating the use of accommodations for instruction and assessment*. Retrieved from http://www.marylandpublicschools.org/NR/rdonlyres/494C2E72-8C85-47C6-B2C4-E2363B81B026/18532/MD_Accommodations_Manual_21108.pdf
- Massachusetts Department of Education. (n.d.). *Massachusetts state report card 2011*. Retrieved from <http://profiles.doe.mass.edu/staterc/ayp.aspx?fyCode=2011>
- Massachusetts Department of Education. (2011). *Bilingual dictionaries and glossaries authorized for use by English language learners on MCAS tests*. Retrieved from http://www.doe.mass.edu/mcas/testadmin/lep_bilingual_dictionary.pdf
- Michigan Department of Education. (2011). *Michigan statewide assessment selection guidance 2011/2012*. Retrieved from http://www.michigan.gov/documents/mde/Michigan_Statewide_ASG_9_19_11_364777_7.pdf
- Michigan Department of Education. (2012). *2011-2012 Michigan educational assessment program proficiency for statewide*. Retrieved from <https://www.mischooldata.org/DistrictSchoolProfiles/AssessmentResults/Meap/PerformanceSummary/Snapshot.aspx?navstatetext=Student+Information&navstatevalue=%2fDistrictSchoolProfiles%2fStudentInformation%2fStudentSummary.aspx>
- Minnesota Department of Education. (2011). *Procedures manual for Minnesota assessments 2011-2012*. Retrieved from http://www.mnstateassessments.org/resources/Manuals/2011-12_Procedures_Manual.pdf
- Minnesota Department of Education. (2012). *Data for parents and educators 2010-2011*. Retrieved from <http://education.state.mn.us/MDEAnalytics/Reports.jsp>
- Mississippi Department of Education. (n.d.). *Mississippi Assessment and Accountability Reporting System 2010-2011*. Retrieved from <http://orsap.mde.k12.ms.us/MAARS/index.jsp>

- Mississippi Department of Education. (2011). *Mississippi testing accommodations manual*. Retrieved from <http://www.mde.k12.ms.us/docs/student-assessment/mississippi-20testing-20accommodations-20manual-20august-2020110784E802769C.pdf?sfvrsn=2>
- Missouri Department of Education. (n.d.). *Accommodations list for students who are English language learners (ELL)*. Retrieved from http://dese.mo.gov/divimprove/assess/documents/Grade_Level_Accommodations.pdf
- Missouri Department of Education. (n.d.). *Missouri accommodations manual*. Retrieved from <http://dese.mo.gov/divimprove/assess/tech/accommodationsmanualpowerpoint.pdf>
- Mitchell, D., Destino, T., and Karam, R. (1997). *Evaluation of English language development programs in Santa Ana Unified School District: A report on data system reliability and statistical modeling of program impacts*. University of California, Riverside: California Educational Research Cooperative.
- Montana Office of Public Instruction. (n.d.). *Adequate yearly progress status information 2010-2011*. Retrieved from <http://opi.mt.gov/Reports&Data/nclb-reports.php>
- Montana Office of Public Instruction. (2007). *Level the playing field with accommodations*. Retrieved from http://opi.mt.gov/pub/PDF/Assessment/Archive/Phase2/Accomodations/08CRT_tentcards.pdf
- Myers, D. E., and Milne, A. M. (1988). Effects of home language and primary language on mathematics achievement: A model and results for secondary analysis. In R. R. Cocking and J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 259-293). Hillsdale, NJ: Erlbaum.
- National Center for Educational Statistics (NCES). (2004). *Language minorities and their educational and labor market indicators—Recent trends* (NCES 2004–09). Washington, DC: U.S. Department of Education.
- National Center for Educational Statistics (NCES). (2008). *Common core of data 2007-2008*. Washington, DC: U.S. Department of Education.
- National Center for Educational Statistics (NCES). (2009). *Common core of data 2008-2009*. Washington, DC: U.S. Department of Education.
- National Center for Education Statistics (NCES) (2012). *National Assessment of Educational Progress (NAEP) 2011 Reading, Math, and Science Results*. Retrieved from <http://nces.ed.gov/nationsreportcard/naepdata/>
- Nebraska Department of Education. (n.d.). *Nebraska state accountability 2010-2011*. Retrieved from <http://reportcard.education.ne.gov/Page/NesaPerfResults.aspx?Category=3&AypGroup=12&Level=st&Subject=1&ScoreID=8>
- Nebraska Department of Education. (2011). *Guide for including and accommodating English language learners (ELLs) in the Nebraska state accountability (NeSA) tests 2011-2012*. Retrieved from http://www.education.ne.gov/assessment/pdfs/2011_12_Guide_for_Including_and_Accommodating_English_Language_Learners_Jan_2012.pdf

- Nesi, H., and Meara, P. (1991). How using dictionaries affects performance in multiple-choice EFL tests. *Reading in a Foreign Language*, 8 (1), 631-643.
- Nevada Department of Education. (n.d.). *Assessment details: CRT state of Nevada 10-11*. Retrieved from http://nde.doe.nv.gov/Assessment_NV_ReportCard.htm
- Nevada Department of Education. (2012). *LEP accommodations form*. Retrieved from http://nde.doe.nv.gov/SpecialEdResources/LEP_Accommodations_Form.doc
- New Mexico Public Education Department. (2003). *Frequently asked questions about English language learner student assessment and accommodations*. Retrieved from <http://www.ped.state.nm.us/BilingualMulticultural/dl08/ELL-Accommodations-FAQs.pdf>
- New Mexico Public Education Department. (2012). *Webfiles proficiencies by grade 2011-2012*. Retrieved from <http://www.ped.state.nm.us/AssessmentAccountability/AcademicGrowth/NMSBA.html>
- New York State Education Department. (2012). *The New York state report card: Accountability and overview report 2010-2011*. Retrieved from <https://reportcards.nysed.gov/statewide/2011statewideAOR.pdf>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002). Retrieved from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_bills&docid=f:h1enr.txt.pdf
- North Carolina Department of Education. (2010). *Guidelines for testing students identified as limited English proficient*. Retrieved from <http://www.ncpublicschools.org/docs/accountability/policyoperations/lep/lepguide1011.pdf>
- North Dakota Department of Public Instruction. (2011). *North Dakota state profile 2010-2011*. Retrieved from <http://www.dpi.state.nd.us/dpi/reports/Profile/1011/ProfileDistrict/99999.pdf>
- North Dakota Department of Public Instruction. (2012). *English learners in North Dakota: Programming*. Retrieved from <http://www.dpi.state.nd.us/bilingul/tech/programming.pdf>
- Ohio Department of Education. (2011). *Ohio statewide assessment program rules book*. Retrieved from <http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=9&ContentID=43712>
- Ohio Department of Education. (2012). *Ohio state proficiency levels 2010-2011 with students disaggregated by limited English proficiency*. Retrieved from http://ilrc.ode.state.oh.us/Power_Users.asp
- Oklahoma State Department of Education. (n.d.). *Annual report card 2010-2011*. Retrieved from <http://www.ok.gov/sde/sites/ok.gov.sde/files/StateReportCard11.pdf>
- Oregon Department of Education. (2011). *Accommodations manual 2011-2012: How to select, administer, and evaluate accommodations for Oregon's statewide assessments*. Retrieved from http://www.ode.state.or.us/teachlearn/testing/admin/alt/ea/accman-update-7-7-2011-12_hc.pdf

- Oregon Department of Education. (2011). *Statewide report card: An annual report to the legislature on Oregon public schools 2010-2011*. Retrieved from <http://www.ode.state.or.us/search/page/?id=1821>
- Pennsylvania Department of Education. (n.d.). *State report card 2010-2011*. Retrieved from <http://paayp.emetric.net/Content/reportcards/RC11M.PDF>
- Pennsylvania Department of Education. (2012). *Accommodations guidelines for English language learners*. Retrieved from http://www.portal.state.pa.us/portal/http://www.portal.state.pa.us;80/portal/server.pt/gateway/PTARGS_0_190926_1241134_0_0_18/PSSA%20Accommodations%20for%20ELs%202012.pdf
- Rhode Island Department of Education. (n.d.). *New England common assessment program accommodations guide*. Retrieved from http://www.ride.ri.gov/assessment/DOCS/NECAP/Test_Admin/NECAP_Accommodations_Guide.pdf
- Rhode Island Department of Education. (2012). *Rhode Island state report card 2011-2012*. Retrieved from <http://www.eride.ri.gov/reportcard/11/State.aspx>
- Rivera, C., Collum, E. Willner, L. S., and Sia, Jr., K. S. (2006). An analysis of state assessment policies regarding the accommodation of English language learners. In Rivera, C., and Collum, E. (Eds). *State assessment policy and practice for English language learners: A national perspective* (pp. 1-173). Mahway, NJ: Lawrence Erlbaum Associates, Inc.
- Rivera, C., and Stansfield, C. W. (1998). Leveling the playing field for English language learners: Increasing participation in state and local assessments through accommodations. In R. Brandt (Ed.). *Assessing Student Learning: New Rules, New Realities* (pp. 65– 92). Arlington, VA: Educational Research Service.
- Rivera, C., and Stansfield, C. W. (2003). The effect of linguistic simplification of science test items on score comparability. *Educational Assessment*, 9 (3/4), 79– 105.
- Sable, J., and Garofano, A. (2007). *Public Elementary and Secondary School Student Enrollment, High School Completions, and Staff From the Common Core of Data: School Year 2005–06* (NCES 2007-352). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Shepard, L., Taylor, G., and Betebenner, D. (1998). *Inclusion of limited-English-proficient students in Rhode Island's Grade 4 mathematics performance assessment* (CSE Tech. Rep. No. 486). Los Angeles, CA: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.
- South Carolina Department of Education. (n.d.). *Palmetto assessment of state standards (PASS) test scores 2010-2011*. Retrieved from http://ed.sc.gov/data/pass/2011/show_state_pass_scores_demo.cfm?ID=999999
- South Carolina Department of Education. (2011). *Accommodations manual: A guide to selecting, administering, and evaluating the use of accommodations for students with disabilities*. Retrieved from <http://ed.sc.gov/agency/programs-services/172/documents/AccomManualMarch2012.pdf>

- South Dakota Department of Education. (n.d.). *Title III: English language acquisition*. Retrieved from <http://doe.sd.gov/oess/TitleIIIela.aspx>
- South Dakota Department of Education. (2011). *No child left behind report card 2010-2011*. Retrieved from <http://doe.sd.gov/NCLB/reports/2011/reportcard/2011state.pdf>
- State of Tennessee Department of Education. (n.d.). *ELL students and Tennessee assessments: 2011-2012 English language learner (ELL) accommodations*. Retrieved from http://www.tn.gov/education/assessment/doc/ELL_Accommodations.pdf
- Stevens, R., Butler, F. A., and Castellon-Wellington, M. (2000). *Academic language and content assessment: measuring the progress of English language learners (ELLs)* (CSE Report No. 552). Los Angeles, CA: University of California, National Center for Research for Evaluation, Standards, and Student Testing.
- Texas Department of Education. (2012). *Texas 2010-2011 NCLB report card*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/nclb/2011/state.html>
- Title I – Improving the Academic Achievement of the Disadvantaged, 71 Fed. Reg. 54188 (Sept. 13, 2006) (34 C.F.R. pt. 200). Retrieved from <http://www.ed.gov/legislation/FedRegister/finrule/2006-3/091306a.pdf>
- U. S. Department of Education. (2003). *Standards and assessments non-regulatory draft guidance*. Retrieved from www.ed.gov/policy/elsec/guid/saaguidance03.doc
- Utah State Office of Education. (2011). *Utah assessment participation and accommodations policy*. Retrieved from http://www.schools.utah.gov/sars/DOCS/assessment/Special_Needs_Accommodations_Policy-pdf.aspx
- Vermont Department of Education. (2010). *English language learners in Vermont: Distinguishing language difference from disability*. Retrieved from http://education.vermont.gov/new/pdfdoc/pgm_esl/educ_ell_sped_resource_guide.pdf
- Vermont Department of Education. (2011). *School accountability system based on student performance: 2010-2011 adequate yearly progress report*. Retrieved from <http://edw.vermont.gov/REPORTSERVER/Pages/ReportViewer.aspx?%2fPublic%2fAssessment+Report+by+Grade>
- Virginia Department of Education. (n.d.). *Virginia school report card 2010-2011*. Retrieved from <https://p1pe.doe.virginia.gov/reportcard/report.do?division=All&schoolName=All>
- Virginia Department of Education. (2011). *Limited English proficient students: Guidelines for participation in the Virginia assessment program*. Retrieved from http://www.doe.virginia.gov/testing/participation/lep_guidelines.pdf
- Washington Office of Superintendent of Public Instruction. (n.d.). *Washington state report card 2010-2011 adequate yearly progress proficiency detail*. Retrieved from <http://reportcard.ospi.k12.wa.us/AYPProficiencyDetail.aspx?groupLevel=District&schoolId=1&reportLevel=State&orgLinkId=1&yrs=16&year=16&chartType=1>

- Washington Office of Superintendent of Public Instruction Washington. (2011). *Washington state accommodations guidelines for statewide assessments*. Retrieved from <http://www.k12.wa.us/Assessment/AlternativeAssessment/pubdocs/AccommodationManual.pdf>
- West Virginia Department of Education. (2011). *West Virginia educational standards test 2 confidential summary report 2010-2011*. Retrieved from <http://wvde.state.wv.us/oa/pdf/results/westest/state/2011/CSR.pdf>
- West Virginia Department of Education. (2011). *West Virginia Guidelines for Participation in State Assessments, 2011-2012*. Retrieved from <http://wvde.state.wv.us/osp/GUIDELINES-TO-PARTICIPATE-2011-2012.pdf>
- Wisconsin Department of Public Instruction. (2012). *The assessment accommodations matrix for English language learners*. Retrieved from <http://dpi.wi.gov/oea/pdf/accomell.pdf>
- Wisconsin Department of Public Instruction. (2012). *Wisconsin school performance report 2010-2011*. Retrieved from <http://www2.dpi.state.wi.us/wsas/StateWkce.asp>
- Wyoming Department of Education. (2006). *Wyoming accommodations manual for instruction and assessment: How to select, administer, and evaluate use of accommodations for instruction and assessment of students with disabilities*. Retrieved from http://edu.wyoming.gov/sf-docs/publications/PAWS_Accommodations_Manual.pdf?sfvrsn=0
- Wyoming Department of Education. (2012). *Public PAWS results state level disaggregated 2010-2011*. Retrieved from <http://fusion.edu.wyoming.gov/Login/web/Pages/PAWSPublicReports/PAWSPublicStateLevelSubgroups.aspx>
- Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., and Stone, E. (2008). Validity and Fairness of State Standards-Based Assessments for English Language Learners. *Educational Assessment*, 13 (2), 170–192.

APPENDIX
STUDENT/TEST ADMINISTRATOR SURVEY

Items for the Student to Complete

Home Language

- English
- Spanish
- Other

Gender

- Female
- Male

Ethnicity/Race

Ethnicity (Mark only one)

- Hispanic or Latino
- Not Hispanic or Latino

Race (Mark one or more)

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White

Items for the Test Administrator to Complete

Special Student Classification

(Mark any that apply)

- ELL/LEP
- F/RL
- Migrant

Test Accommodations

(Mark any that apply)

- Read-aloud
- Tested Off Level
- Extended Time
- Provision of English/Native Language Word-to-Word Dictionary (no definitions)
- Test Administered by ELL Teacher or Individual Providing Language Services
- Repeated Directions
- Other