Addressing the Inclusion of English Language Learners in the Educational Accountability System: Lessons Learned from Peer Review

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Laurene L. Christensen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

R. Michael Paige, Ph.D., Adviser

January, 2010

ACKNOWLEDGMENTS

It is not the mountain we conquer but ourselves.
~Edmund Hillary

The fall semester of 2009 marks the twentieth anniversary of my starting college. My parents drove me six hours from home and dropped me off with a set of new clothes, a new backpack, big hair and big earrings, and big aspirations for my future. I don't know that my education has happened in a way that anyone could have predicted. I, for one, know that I have learned more than I ever thought I could, and as many or more of those lessons have come from listening to people, as they have from reading books.

A friend of mine once told me that education is about the journey, not the destination. In some ways, finishing this dissertation feels a little like I've reached the destination. It's been a long journey! I couldn't do it alone. There are many people I want to thank, who have helped me along this journey:

First and foremost, I want to thank my committee. Special thanks go to my adviser Dr. Michael Paige. I can't thank you enough for all of your help and encouragement. I especially have appreciated our good talks, and two-hour advising sessions! Thanks again for co-teaching with me. Thanks also to Dr. Constance Walker for chairing my dissertation committee, and for sharing your perspective on English language learners and assessment. I also enjoyed co-teaching with you. To Dr.Walt Jacob, my thanks and appreciation for sticking with me through the dissertation process, even through a dramatic topic change. The next time I see you at Target, you won't have to ask me how my dissertation is going! Finally, I want to thank Dr. Kimberley Brown for all of her years of support and encouragement, for great conversations on the future of language education—which most likely includes assessment—and above all, for pushing me in the direction of a few major life decisions.

Along my education journey, there have always been important women who have served as mentors and role models for me. Each of you has contributed in your way, edging me along to the next phase of the journey. My high school English teacher, Judy Kraft, encouraged me to dream big, and I have carried that message with me. Sharon Carson inspired my confidence to continue on to further education beyond my undergraduate degree. Susan Danielson helped nurture my thinking to a better understanding of social justice. I have appreciated the support and friendship of each of you over the years.

At the University of Minnesota, I have had the wonderful experience to have a great group of women as a dissertation support group. This process has been long and arduous, and your friendship over the years has made it all that much easier to get this job done. Thorunn Bjarnadottir, Holly Emert, and Kaoru Kinoshita—thanks so much for all of your support and encouragement.

I can't say enough how supportive my workplace has been as I have been in the final stages of completing this dissertation. Martha Thurlow has been an amazing mentor and champion of me, and I feel so fortunate to have landed at the National Center on Educational Outcomes. My fabulous co-workers, including Michael Moore, Dorene Scott, Kristi Liu have given me so much encouragement these last few years, and I am grateful to all of you. Other professional colleagues, including Courtney Foster and Chris Bremer have given me ideas and encouragement.

When my educational journey brought me from Portland to Minneapolis, I left behind some incredible friends who continue to cheer me on. Heather Gaddy Blackbird saw this day coming all along, even when I couldn't believe it. Carol Burnell has continued to inspire me to write, write, write. Kate Sage and Debra Clemans have both been supportive listeners as

life has intervened in my process, and I appreciate your thoughtful listening. Here in Minnesota, I've cultivated friendships along the way, and these additional supporters, including Dana Lundell, Vitaliy Shyyan, Molly Coyne, Liz Templin, and Jeffrey Schulz have continued to make my journey worthwhile. A special thanks goes to Robin Murie, for always being on my team (and for encouraging me to learn to knit). Words don't really convey the appreciation I have for how each of these individuals has touched my life.

Through it all, my family has been an important part of this journey. We have had ups and downs, no doubt, and right now, I am deeply grateful for all that my family has done to support me over the years: my brother Steve, who helped me make the move to Minnesota; my sister Rae, for helping me keep my priorities in check; my sister Jan, for baking too many goodies for my defense meeting; and my sister Pam, who has had many riveting conversations with me about education policy and the impact of testing on kids. Above all, though, I'd like to thank my parents, Bill and Lola Christensen, for always believing in me, and for knowing just how much teasing is enough to keep me going. Now you can say there's a doctor in the family.

Many nights of writing into the wee hours, I have been supported by little Java, our sweet dog who did not live to see this project come to an end. But nonetheless, he had a paw in this, and I will always remember his love and devotion. His companion, Sappho, and his successor, Pearl the cat, have taken over his watchful duty, and I have enjoyed their companionship. Writing can be a lonely task, and the quiet friendship of animals is a gentle reminder that I am not alone.

When I began my postsecondary education with aspirations toward an advanced degree, I did not foresee having children along the way, but

someone wiser than I suggested that children make a person a little less selfish, and that persuasion was enough. I am truly blessed with two wonderful children, Jurgen and Sigrid, and to say that they have been a motivating force for me would be an understatement. Having children may have slowed things down a bit for me, but the pace of this part of the journey has been exactly right for me to enjoy every part of it and has made getting this degree far more meaningful than I ever imagined it could be.

Finally, through it all, I have had the steadfast love, support, and companionship of Christopher Rogers, my spouse. Words cannot convey the appreciation I have for all he has done in walking with me in this journey together.

This thesis is dedicated to Lola Christensen,
for starting me out on this journey and seeing me through to the end,
and
to Sigrid Rogers,
for being the inspiration I needed at just the right time.

ABSTRACT

This study investigated the inclusion of English language learners (ELLs) in state standards and assessments, as measured by comments made by peer reviewers in the federal evaluation of states' standards and assessments. As required by the Elementary and Secondary Education Act (ESEA), reauthorized in 2004 as No Child Left Behind (NCLB), states are required to develop challenging content standards in English language arts and mathematics, and to assess students on these standards. Nationwide, English language learners have been performing below proficient on these state tests. The goal of this study was to determine the elements of the standards and assessment system that peer reviewers thought could be addressed in order to improve the accessibility of these assessments for ELLs.

This dissertation employed two research methods: a frequency and distribution analysis of the peer comments relevant to ELLs across the seven Critical Elements of the peer review, and a thematic analysis of the comments, using inductive methods to establish themes.

The study found that comments relevant to assessing English language learners were present in all seven critical areas and in all states. All states had two or more critical elements with comments relevant to ELLs. Key themes from the qualitative analysis included accommodations, diverse stakeholders, reporting accuracy, inclusion, reporting for parents, migrant students, comparability of test forms, bias review, DIF analysis, and standards. States have begun to address the inclusion of ELLs in statewide standards and assessments; however, state agency personnel have more work to do in these areas before assessments can be seen to be fully accessible to ELLs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

With realization of one's own potential and self-confidence in one's ability, one can build a better world. –Dalai Lama

As a student, I have often enjoyed taking tests. In elementary school, we took the Iowa Basic tests each year, and I looked forward to a day of quiet testing and Number 2 pencils. The smiling faces of the children on the test booklet were my own, as I happily darkened the ovals on the test answer page. The satisfaction I felt on test day was reinforced when my mother reported my results at home. I had done well, she affirmed, trying not to let my confidence spoil me.

I have always known that not everyone has had the positive experiences with testing that I have had. My best friend in high school struggled with test anxiety, and even though we prepared for the ACT test for six weeks in my high school English class, her final score did not truly represent her knowledge and skills. In graduate school, I have had several international student friends who have felt psychologically defeated by the TOEFL and GRE exams. These tests are used as gatekeepers, and in some cases, just their very existence is enough.

Some of my most compelling experiences with testing have been with the high school students I had the opportunity to teach through the Commanding English program—a high school to college bridge program—at

Edison and Washburn High Schools in Minneapolis, Minnesota. In these schools, I taught gifted and ambitious English language learners who had aspirations for college. Together, we did some preparation for the ACT not unlike what my high school English teacher had done with me. Yet, I found myself explaining cultural references that were taken for granted by me, and simultaneously unheard of by my students.

Other tests posed challenges for my students as well. One student, "Safiya," won a scholarship for students who had high academic achievement and had also overcome great hardship in life. Safiya had spent time in a refugee camp in Kenya before coming to Minnesota, and she lived in the state with no family, supporting herself through school. However, Safiya was not able to immediately use her scholarship to go to college because she had failed her Minnesota Basic Skills Test in Writing, and this was required for graduation. Having read Safiya's writing, I know that the test did not adequately capture Safiya's skill with the subject.

Nonetheless, I believe that these assessments are important. Other stories my students shared with me, and stories from others, confirm that ELLs do not always have full access to grade level curriculum. Sometimes this happens in spite of educators' best intentions. We might hear a compelling story from a student about their survival in a boat on the Mekong Delta and

want to go easy on the student, rather than hold them to high expectations. But, when the student loses the opportunity to learn, we all lose. Currently, assessments are the tool we have for ensuring that all students have access to the curriculum. These assessments can use some improvement, so that they are more inclusive of ELLs. When students like mine take these tests, they should believe that their score will be an accurate representation of their knowledge and skills. They should feel confident, like a wide-eyed girl in western North Dakota once did, when they bubble in their answers or answer with a click of a mouse. I hope this thesis is a small step in that direction.

CHAPTER I: INTRODUCTION

Mountains cannot be surmounted except by winding paths.
--Johann Wolfgang von Goethe

Every year, in schools across the United States, kids take tests. They take different types of tests, and for different reasons. Sometimes a test might be for placement; other times a test determines mastery. A test may be given at several different points to monitor progress. A test may be developed by a teacher or by a testing company. Tests may be developed by state departments of education. These tests are used to determine whether or not public schools are doing what they should to make sure all students, including English language learners and students with disabilities, are held to the high academic standards developed by that state.

This research study is about how state departments of education address the academic achievement of students in public schools in the United States who speak a language other than English at home. These students—usually called English language learners, sometimes also referred to as limited English proficient students—are a group of students whose presence in US schools has steadily been increasing over the past 10 years (Wolf, Herman, Bachman, Bailey, and Griffin, 2008). For the 2005-06 school year, states

reported approximately 5 million ELLs enrolled in U.S. schools, according to information provided by the Office of English Language Acquisition (OELA). Current estimates suggest that English language learners are the fastest growing subgroup in US schools (Office of English Language Acquisition, reported by Wolf, et al., 2008). English language learners also represent a diverse group of students. As a group, they speak over 400 different languages (Wolf, et al., 2008). They include students with and without disabilities. They have a range of previous schooling experiences. They may or may not be migrant students. As a group, however, "[t]he sizable ELL population typically fails to meet the proficient level in academic standards, and the academic gap between this group and the non-ELL population is considerable" (Wolf, et al., p. 2).

The US Department of Education has been concerned for some time about the achievement of ELLs, and intends to ensure that their educational needs are increasingly being addressed by states' assessment and accountability systems. The reporting of results from statewide content assessments are one means by which the federal government ensures that students are learning what the state standards intend for students to learn. This study seeks to understand the key issues that have been raised with regard to including English language learners in the statewide content

assessments, i.e., English language arts and mathematics, required by the accountability process.

The purpose of this study is to highlight the key issues regarding the inclusion of English language learners in the accountability system. The study draws on data from the current federal peer review of state standards and assessments. Knowing the issues relevant to English language learners from the peer review will be important to the field. State agency personnel want to have more inclusive assessment systems, and they need help identifying the issues that need to be addressed in order to make their systems more accessible. Researchers and other experts concerned about English language learners will be able to use information from this study in order to prioritize research and policy agendas relevant to the needs of ELLs. Addressing the needs of English language learners in the assessment system is one of increasing importance.

This chapter introduces the context relevant to both the assessment of English language learners and the standards and assessment peer review, raises the research question central to the current study, describes the sources of data that will be used, and provides a glossary of relevant terminology.

Background of the Study

The current United States elementary and secondary school context is one strongly influenced by accountability. The most recent reauthorization of the Elementary and Secondary Schools Act of 1965 (ESEA) increased accountability requirements for states, districts, and schools in order to receive federal funding. Commonly known as No Child Left Behind (NCLB), this legislation required states to implement standards-based reform, which has also been sometimes referred to as outcomes-based reform. A key assumption of this model is that high expectations, along with quality instruction, will result in high achievement for *all* students. The popular name for the law, No Child Left Behind, refers specifically to the expectation that by 2014, all students will be scoring at or above proficient on state accountability tests. NCLB requires states to assess students in reading, math, and science and to publicly report the results of those assessments. Schools that fail to meet their goals for adequate yearly progress (AYP) may lose federal funding if they fail to make improvements within five years (Public Law 107-110).

NCLB represents a growing concern at the federal level about the educational achievement of students in special populations, such as students with disabilities, students from racial/ethnic minority backgrounds, and students who speak a home language other than English. NCLB has special

provisions to address the achievement of each of these groups; this study focuses on English language learners.

Who are English Language Learners?

Students who speak a home language other than English may be referred to as Limited English Proficient (LEP) and/or English language learners (ELL). LEP is a term used primarily by the federal government and state education agencies to describe students whose English proficiency level precludes them from participating meaningfully in general education classrooms. An LEP student was defined, in an earlier reauthorization of the Elementary and Secondary Education Act (1994) as one who,

> has sufficient difficulty speaking, reading, writing, or understanding the English language and whose difficulties may deny such individual the opportunity to learn successfully in classrooms where the language of instruction is English or to participate fully in our society due to one or more of the following reasons:
>
> - was not born in the United States or whose native language is a language other than English and comes from an environment where a language other than English is dominant;
> - is a native American or Alaska native or who is a native resident of the outlying areas and comes from an environment where a language other

than English has had significant impact on such individual's level of English language proficiency; or

- is migratory and whose native language is other than English and comes from an environment where a language other than English is dominant (sec. 7501).

This federal definition may or may not be used by states. Historically, states have held their own definitions of LEP as well as made determinations about how students qualify to enter and exit from English language support programs (Anstrom, 1996).

The term English language learner (ELL), sometimes also called English learner (EL), is a broad term that may refer not only to those students who have limited English proficiency but also to students who have exited from English language support programs or students who are bilingual and have not received English language support services. The use of the term ELL, or EL, is growing in preference over LEP, in part, because LEP frames the student's English proficiency level as a limitation rather than focusing on the fact that the student is multilingual and is learning English. However, federal legislation continues to use the term LEP. Thus, in this research study LEP is used in reference to the term's usage in federal law. However, ELL is used whenever possible.

English Language Learners in U.S. Schools

According to the National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, the enrollment of students who have limited English proficiency (LEP) in the 2005-06 school year was just over 5 million students. This number represents an increase of 57.17% from the 1995-96 school year. Roughly 1 in 10 students in United States elementary and secondary public schools is an English language learner. Provisions for English language learners are addressed in No Child Left Behind under both Title I and Title III of the Act.

Title I of NCLB addresses the academic achievement of students who are seen to be in disadvantaged groups, including the following students:

(1) low-achieving children who are in impoverished schools,

(2) limited English proficient children (LEP),

(3) migratory children,

(4) children with disabilities,

(5) Indian children,

(6) neglected or delinquent children, and

(7) young children in need of reading assistance (Sec. 1001).

Under Title I, states are expected to develop standards that address the academic achievement of the student groups listed above. In addition, states must also develop assessments in reading and math, and science beginning in

2008, that are aligned to these standards. Students must be administered the assessments yearly in grades 3-8, and once in high school.

With regard to English language learners being assessed under Title I, schools may exempt English language learners from English language arts assessments the first year; however, English language learners may not be exempted from math assessments. In addition, Title I requires that appropriate accommodations be given to English language learners. Accommodations, such as a word-to-word glossary or extra time, allow students to access the assessment and "level the playing field;" accommodations do not reduce learning expectations for students.

Title III of No Child Left Behind--also referred to as the English Language Acquisition Language Enhancement, and Academic Achievement Act—addresses language instruction for students who are seen to be limited English proficient (LEP) students. Under Title III, states are expected to help children to develop proficiency in English and also to meet "the same challenging State academic content and student academic achievement standards as all children are expected to meet…" (Section 3211). Thus, states must assess a student's level of proficiency in English in addition to expecting the student to participate in the same content assessments as native-English-speaking students.

Since its approval by Congress in 2001, NCLB has been met with mixed reviews. Supporters, including many disability advocates and parent advocates, praise the law for increasing access to the general curriculum for students with disabilities and other disadvantaged student groups. However, the law's detractors often note that NCLB has not been fully funded, and state and local education agencies need financial support in order to enact the reforms the law mandates.

The purpose of this research study is not to critique the law's merits and limitations. Rather, this research study proposes to examine the peer review process implemented as part of accountability requirements under Title I of NCLB to discover what states need to do to improve their English language arts, mathematics, and science assessments to make them more accessible to English language learners.

Overview of the Standards and Assessments Peer Review Process

As part of the accountability components of NCLB, states must demonstrate to the federal government that they have enacted content standards and have developed and implemented assessments aligned to these standards. To ensure that states meet these requirements of the law, the U.S. Department of Education (ED) has been regularly using a peer review process to ensure that states are in compliance with components of the Elementary

and Secondary Education Act. While the topic of each peer review may change, the process itself is used regularly by the Department for decision-making. The current peer review process, begun in 2004, is on the topic of standards and assessment. That is, experts in the field of standards and assessment have been reviewing the evidence compiled by States to demonstrate that each State's assessment system holds schools and school districts accountable under NCLB. The experts, referred to in the study as "peers," include a three-person team: an expert in psychometrics, a representative from a state department of education, and an expert in special populations, usually special education or English language learners but not both. The peer review teams vary by state, and the specific membership of each review team is not known to the state under review.

The peer review process began by evaluating each state's mathematics and English language arts standards and assessments. Since then, the evaluation has expanded to include alternate assessments based on modified academic achievement standards and also science assessments. Several of the criteria in the evaluation focused on issues relevant to English language learners.

At the start of the peer review process, each State compiled a set of evidence materials, including state statutes and regulations, test administrator

manuals, board resolutions, and assessment reports (U.S. Department of Education, 2004, p. 6), to convince the peer reviewers, a team of experts on assessment and standards, that the State assessment system meets NCLB requirements. The reviewers, under the guidance of an ED staff person, have been reviewing the materials to determine that the state assessment system complies with NCLB requirements. The reviewers provided feedback to states that includes changes required to meet NCLB requirements as well as guidance for improving the state's assessment system. Many states have received approval notification; other states have been asked to provide additional evidence.

The peer review of standards and assessments considers seven broad sections, which ensure that the state's system has the following elements:

(1) A single statewide system of challenging academic content standards applied to all public schools and LEAs;

(2) A single statewide system of challenging academic achievement standards applied to all public schools and local education agencies;

(3) A single statewide system of annual high-quality assessments;

(4) A system of assessments with high technical quality;

(5) Alignment of academic content standards, academic

achievement standards, and assessments;

(6) Inclusion of all students in the assessment system; and

(7) An effective system of assessment reports (U.S. Department of Education, 2004, pp. i-ii).

Within each of these sections, there are critical elements that directly relate to English language learners. These sections and their implications for English language learners are explained in more detail in Chapter 2.

Each section of the peer review guidance is broken into a total of thirty-eight sub-units referred to as "Critical Elements;" thus, each section has approximately 5-6 critical elements. For each Critical Element, the peer review guidance includes a description of the element, followed by a description of examples of Acceptable Evidence and Incomplete Evidence. Please refer to Appendix A for sample pages from the Standards and Assessments Peer Review Guidance: Information and evidence for Meeting Requirements of the No Child Left Behind Act of 2001.

The peer review serves two primary purposes: (1) to provide feedback to assist states in making improvements to their assessment systems, and (2) to serve the Assistant Secretary for Elementary and Secondary Education in making decisions about the approval of each state's assessment system. States that do not meet the requirements of NCLB may have their federal Title

I funding withheld. Title III funding is given or withheld through a different process.

In early rounds of the peer review process, several states did not provide sufficient evidence to demonstrate the full inclusion of English language learners in the assessment accountability system, and thus were not approved. States were then asked to provide additional information in order to gain approval. For these states, continuing to try to gain approval has become challenging because state education agency personnel may be uncertain about what additional evidence to submit (Christensen, Lail, and Thurlow, 2007; Thurlow, Christensen, and Lail, 2008). When other areas of the peer review process have been examined (Christensen et al., 2007; Thurlo et al., 2008), states have found this additional information to be very helpful to their own process of submitting additional information. Having a big picture understanding of the peer review process, as well as learning from other states' successes can be useful for states that continue to have outstanding peer review issues. Furthermore, this aggregated view of issues that emerged from states' peer reviews with regard to ELLs can help the field prioritize research and policy decisions for ELLs.

Purpose and Significance of the Study

Donovan, Bransford, and Pellegrino (1999) have described the bidirectional effect that educational research and classroom practice have on each other. Both educational research and classroom practice are influenced by and are able to influence the development of curriculum, teacher education programs, educational policy, and public perceptions (see Figure 1.1). The study undertaken here involves both policy and educational materials. The Standards and Assessment peer review is an evaluation of policy implementation, and the standards and assessments themselves are educational materials. As a result, this thesis has significance for educational policy and the development of assessments. In addition, the results of this study are significant for teacher education, classroom practice, and public perception.

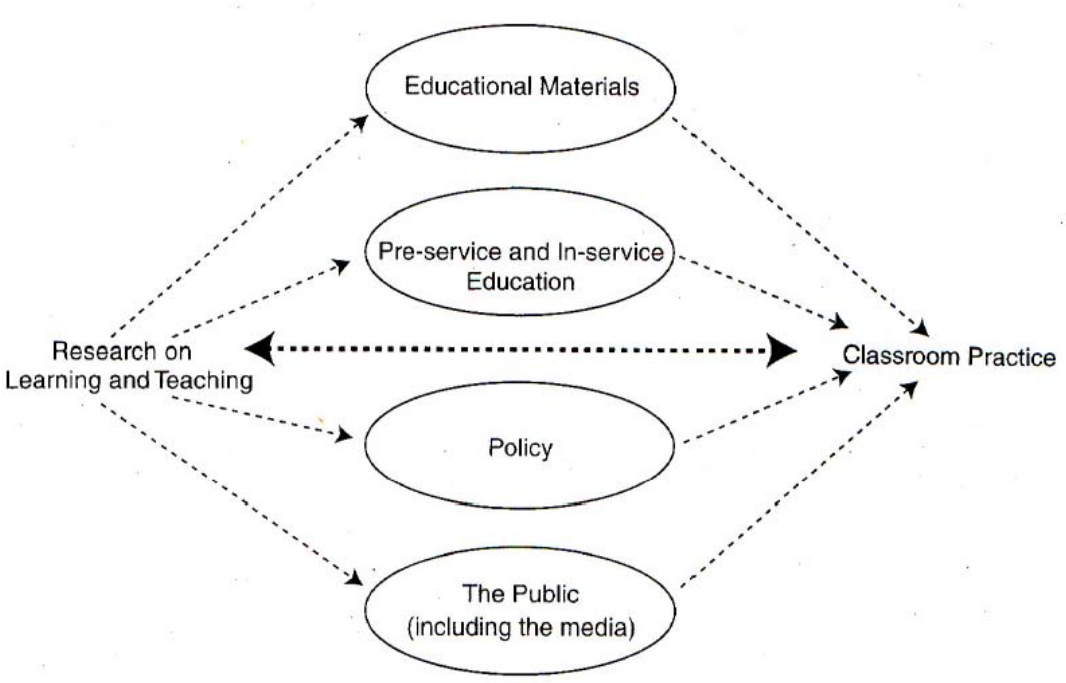

Figure 1.1 Paths through which research influences practice (Donovanet al., 1999, p. 7)

As a study of the implementation of components of NCLB, this thesis has implications for policy. States receive federal monies when they are in compliance with the requirements of NCLB, and this money is withheld when states fail to meet the law's expectations. The peer review process began in 2004, and by the end of 2006, only eighteen states had received full approval of their assessment systems. For many states, the process of gaining approval has been challenging. In my own conversations with state education agency personnel, I have learned that often times, state personnel are unsure of what evidence to submit in response to a particular outstanding critical element.

The purpose of this study is to focus on the elements of the standards and assessments peer review that are relevant to English language learners, and to draw attention to the elements that states are doing well in this regard as well as the elements that state departments of education need to further address. In this way, those who develop assessments at the state education agency level will benefit by knowing those areas to address in improving assessments, the educational materials represented in Figure 1.1, for ELLs.

This study has the potential to influence teacher education and classroom practice. By knowing the key areas of concern in the development of assessments for ELLs, teachers and teacher educators can make more informed decisions in their classrooms. For example, by knowing that accommodations for ELLs is an area in which assessments and assessment policies are lacking, teacher educators may spend more time helping pre-service teachers learn about accommodations decision-making.

Through this study, researchers and policy makers who are concerned with issues relevant to ELLs may have a better idea of which issues are most challenging for states with regard to the accountability system. The study itself may be seen to be a small-scale test of the accountability system. That is, given that ELLs generally score below proficient on accountability assessments, one might expect there to be several outstanding issues in state

assessment systems related to including ELLs. The accountability system has been developed on the basis of closing the achievement gap, and state education agency personnel and test developers are still grappling with how to include students like ELLs, who had previously been excluded from statewide assessments. The issues identified through this study may be key areas to address in order to better include ELLs in assessments. In some cases, more research in these areas may be needed in order to best address the inclusion of ELLs.

The last path of influence described in Figure 1.1 is the public/the media. It is my hope that state departments of education will be able to use the information from this research study to improve their accountability systems—their standards and assessments—so that they may be more inclusive of English language learners. When state assessments are more inclusive, ELLs will be better able to show what they know and can do, and the reporting of assessment results will be more positively received by the public. While not on a grand scale, this information may do something to improve the public perception of ELLs in this country.

Research Questions

The central questions this study proposes to answer are the following:

1. What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be in compliance with the standards and assessments accountability requirements of No Child Left Behind?

2. What are the commonalities among states in meeting the federal requirements for assessing English language learners?

3. What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest?

Data for the Study

Data for this study will be drawn from two sources: (1) the peer review comments included as feedback for the states; and (2) the decision-letters mailed to states from the Department of Education. The peer review comments have been released to the researcher with permission from OESE staff. The decision-letters are publicly available on the Department of Education's website, www.ED.gov.

The peer review comments are organized in the same format as the guidance document for the peer review process. These comments are listed anonymously and collected on one document by the US Department of Education staff person who facilitated the peer review process. The comments

point out some positive attributes of a state's standards and assessment system. Some comments also ask the state agency personnel questions, or point out key issues to focus on. Most peer comments observe shortcomings in the state's accountability system. The comments include margin comments that directly relate to the subsections of the guidance as well as summary comments at the end of each of the seven sections. Most of these margin notes directly relate to elements that are missing, in other words, elements that the state needs to address in order to pass the peer review. However, other comments include suggestions for improving the assessment system; these additional comments may be considered recommendations for best practices.

The peer comments will be collected from the beginning of the peer review process, which began in early 2004, through to December 2006. At that point, the peer review process became somewhat more informal, relying on email and phone conversations between state agency personnel and US Department of Education officials. It is worth noting, however, that by December 2006, each state had gone through at least one round of the peer review process and all of the major issues in each state's assessment system had been identified.

The decision letters are less specific than the peer comments. The decision letters let the state know if they have met the requirements of NCLB for high quality standards and assessments. If the state has not met the federal requirements of peer review, the decision letters detail what needs to be done for the state to comply with NCLB. In most cases, states did not pass the peer review process with only one round of review. As a result, most states have two or more letters that indicate major issues that need to be addressed. The decision letters serve as an important element of triangulating the peer comments for two reasons. The first is that, as stated previously, the peer review process has become less formal over time, so the letters themselves serve as an important indication of the key issues that each state must address. In addition, the decision letters demonstrate the relative importance of the issues raised by the peer reviewers. In some cases, there may be several peer comments included for a state's documentation. However, when one considers the major issues as delineated in the decision letters, this issue may be less critical than the comments indicate.

Research Methods

The research method used in this study involves mixed methods. First, a descriptive quantitative analysis will examine the frequency by which ELLs are mentioned in the peer comments for each section of the peer review.

Then, a qualitative analysis of peer comments and decision letters will be carried out. These data will be analyzed using content analysis techniques. Themes will be determined both deductively, using the 7 sections of peer review, and inductively, allowing additional themes to emerge from the data as necessary. As there are two sources of data, a comparison between the two data sources will be used to ensure that all themes are identified. Additional validation will include sharing the themes and evidence for those themes with a US Department of Education staff person who is on the peer review team and has agreed to serve in this capacity for the study. More details on the research methods for this study will be described in detail in Chapter 3.

Definition of Terms

The following items are terms used frequently throughout this research study.

Accommodations are changes to either instruction or assessments, or both, that include presentation, response, settings, scheduling, and use of equipment that allow students with disabilities and English language learners to access instruction and/or assessment without changing the learning expectation for the student. Examples of accommodations for English

language learners might include oral translations, use of a translation dictionary, or testing in an ELL classroom.

Alignment refers to the relationship between the content standards and the items on an assessment.

Alternate assessment is a type of assessment used for students with disabilities who are considered to be unable to take the general assessment. An alternate assessment may be based on alternate achievement standards, grade level achievement standards, or modified academic achievement standards. States are required to provide at least one type of alternate assessment for students with disabilities.

Alternative assessment is a type of assessment in which the presentation of the assessment has been altered but the questions and constructs being measured remain the same. Alternative assessments for ELLs include native language assessments and simplified English assessments.

Assessment is defined by the American Educational Research Association (AREA), American Psychological Association (APA), and National Council on Measurement in Education (NCME)'s (1999) Joint Standards for Educational and Psychological Testing as "any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects or programs" (p. 172). Pellegrino, Chudowsky, and Glaser

(2001) defined assessment as “the means used to measure the outcomes of education and the achievement of students with regard to important competencies” (p. 52). Several types of assessments are commonly used in public education settings, including teacher-designed classroom assessments, off-the-shelf assessments used in the classroom, portfolio-style assessments, high stakes assessments such as graduation tests or college-entrance examinations, and statewide assessments used for accountability. This research study focuses on the latter, statewide assessments.

Bias review is a process by which test developers ensure that test items are fair and equitable to all test takers. A bias review usually involves gathering feedback on test items from a diverse panel of stakeholders.

Comparability refers to the ability to state that scores from one test form are equivalent to scores from another test form. In some cases, these different test forms may be in another language, or they may be in simplified English.

Consequential validity is a type of validity that considers whether the assessment is having the effect intended. Studies of consequential validity seek to uncover both intended and unintended consequences as well as positive and negative effects of the assessment.

Criterion-referenced test is a type of test in which the score is measured against a standard, or criterion, rather than to other test takers. Using a criterion-referenced test, it is possible for all test-takers to score within the proficient range.

Differential Item Functioning Analysis (DIF Analysis) is a process by which test developers determine the fairness of test items for different demographic groups of students.

Inclusion refers to the practice of ensuring that all students are able to receive access to instruction and to participate fully in assessments.

Instruction can be defined as "the methods of teaching and the learning activities used to help students master the content and objectives specified by a curriculum. Instruction encompasses the activities of both teachers and students. It can be carried out by a variety of methods, sequences of activities, and topic orders" (Pellegrino et al., 2001, p. 52).

Limited English Proficient (LEP) is defined by the federal government as a student whose English language proficiency level precludes him or her from participating meaningfully in general education classrooms. LEP students are also sometimes referred to as English language learners.

Migrant student is a student who is employed in seasonal agriculture or who has an immediate family member employed in seasonal agriculture. A migrant student may or may not be an English language learner.

Native language assessment is an alternative form of the general assessment that is provided in a language other than English. A native language assessment may be either a direct word-for-word translation of an assessment, or it may be a dual-language side-by-side translation with English on one side, and the student's native language on the other side.

No Child Left Behind (NCLB) is the current reauthorization of the Elementary and Secondary Education Act of 1965. In its current form, the law requires a high degree of accountability for states, districts, and schools, to ensure high quality education for all children, including English language learners.

Norm-referenced test is a type of test in which the scores of test takers are compared with each other, rather than to a standard. To that end, the scores of test takers are ranked according to percentiles, with some students at the top and some at the bottom. With this type of test, not all students can be seen as proficient.

Participation includes the different ways that English language learners can be included in the assessment system. Many ELLs can take the general

assessment without accommodations. Some ELLs may need to use accommodations, including translated test forms or simplified English forms of a test. Other ELLs may have disabilities and may need additional accommodations. A very small number of ELLs may have a significant cognitive disability and may need to take an alternate assessment based on alternate achievement standards.

Simplified English assessment is an alternative assessment form in which the structural linguistic elements of the language have been changed so that the language load is easier for the student. Changes may include limited use of pronouns and passive verbs, for example.

Title I is the section of the No Child Left Behind act that requires states to establish content and achievement standards and assessments aligned with those standards, as well as to publicly report the results of the assessments. Title I provides financial support to states to ensure that students who may be disadvantaged are given access to high quality education. States that do not comply with the law can lose federal funding.

Title III is the part of the No Child Left Behind Act that requires schools to ensure that students develop proficiency in English. Under Title III, states set Annual Measurable Achievement Outcomes (AMAOs), and can be penalized for failing to meet these targets. The penalties for not meeting Title III outcomes

is not as severe as the penalties under Title I, in part, because the financial support to states under Title III is not as great as it is under Title I.

Title VII is also known as the Bilingual Education Act. This Act was passed in 1968 in order to provide resources to states to support language minority students. This Act was sharply modified within the 1970s and 1980s, and ultimately was reshaped to become support for English as a second language programs.

Organization of the Thesis

This thesis is organized in five major chapters. This introductory chapter has established the background of the problem and introduced the research question. Chapter 2 reviews the literature relevant to the peer review process, including the historical background that led to the accountability movement. Studies on the peer review are considered. Finally, Chapter 2 provides a detailed overview of the current key issues in the assessment of ELLs. Chapter 3 discusses the methodology used for the current study. A general overview of mixed methods is provided, along with a detailed protocol explaining how the data was analyzed. The results, both quantitative and qualitative, are shared in Chapter 4. This chapter begins with the reporting of the quantitative data, including a matrix that shows the frequency in which issues related to ELL assessment occurred. Themes that relate to the seven

sections of the peer review are then discussed, followed by the themes that emerged inductively. Finally, this thesis concludes with discussion in Chapter 5 that includes a discussion of the implications of this study as well as directions for further exploration.

CHAPTER II: LITERATURE REVIEW

The great aim of education is not knowledge, but action.
--Herbert Spencer

In current educational context, the inclusion of English language learners in statewide assessment systems is a significant issue of concern among educators. For example, in the upcoming 2009 National Conference on Student Assessment sponsored by the Council of Chief State School Officers, roughly 12 percent of the conference is devoted to sessions that focus directly on the assessment of English language learners (CCSSO, 2009). States are challenged, on the one hand, to include English language learners in the assessment system because including students means that they have access to grade-level curriculum and that they are represented in the education system's successes and failures. On the other hand, test developers are challenged to create tests that are fair to English language learners; many factors must be considered in order for a test in English to demonstrate what an English language learner knows and can do. Similarly, educational policy makers must establish policies that take into account the inclusion of English learners on statewide tests. NCLB expects that all English learners will demonstrate proficiency on content assessments by 2014. Thus, states are scrambling to address these issues in their assessment systems, and to

demonstrate that they have met these requirements through the peer review process.

This chapter reviews the current literature relevant to this challenge. Two primary topics are reviewed here: (1) literature on the assessment of English language learners and (2) studies that have been conducted on the peer review process. The assessment of English language learners, as well as the peer review process, are both situated within a larger context of educational reform. Thus, models of policy analysis and educational effectiveness are discussed as they relate to the topic under consideration.

Relationship among Assessment, Instruction, and Policy

As mentioned in Chapter I, ELLs are among the fastest growing subgroup of students in US schools. According to information by the General Accounting Office, for example, ELL students' math proficiency level averaged 20% lower than that of the overall population in 2003–2004 across 48 states (reported by Wolf, et al., 2008).At the same time, they are also consistently scoring below proficient on accountability assessments. This performance gap between ELLs and their native English speaking peers has been documented over time by others, including Abedi and Gándara (2006); Abedi, Herman, Courtney, Leon, and Kao (2004); and Abedi (1998), among others. In response to this observed lower assessment performance on the part of ELLs as a subgroup, educators and policymakers may wonder if this phenomenon

is due to lack of access to instruction, an assessment that does not allow ELLs to show what they know and can do, some other factor, or some combination of all of these.

One primary goal of assessment is to observe what students know and are able to do. About assessment, Pellegrino et al. (2001) wrote:

> Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students' performance, and an interpretation method for drawing inferences from the performance evidence thus obtained. (p. 2)

This relationship among student knowledge, instruction, and assessment has been described as a triangle, known as the "assessment triangle" (Pellegrino et al., 2001), which is pictured in Figure 2.1 below.

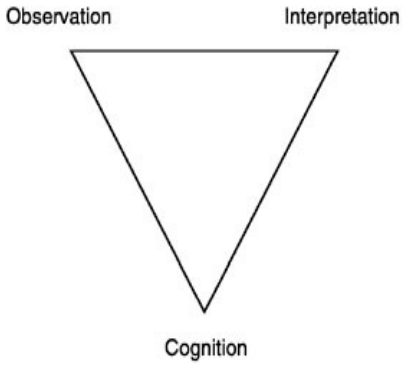

Figure 2.1 The Assessment Triangle

The Assessment Triangle is built on the notion of evidence-centered assessment development. Thus, in designing an assessment, a test developer should ask, "what does competency in this area look like?" That is, what evidence does a student need to produce to demonstrate that they are proficient in a certain skill or competency? Thus, the triangle rests on *cognition*—what does the student know that needs to be assessed? *Observation*, in the upper left corner, refers to the tasks that, when performed by the student, will demonstrate the evidence needed. *Interpretation*, in the upper right corner, refers to the methods and tools used to draw conclusions about the observations that have been collected (Pellegrino et al., 2001, pp. 44-49). These methods may vary, depending on the type of assessment. For example, a classroom teacher may use a rubric; a large-scale assessment may require statistical modeling.

Each corner of the triangle is interrelated with the other corners. Thus, in considering the performance of ELLs on accountability assessments, it may be the case that the tasks are not adequately designed so that the student can demonstrate his or her cognition. Similarly, the interpretive methods may not appropriately account for ELLs as a subgroup.

As Pellegrino et al. (2001) asserted, "assessment is not an isolated part of the education system" (p. 51). What is measured on an assessment and

how stakeholders use the information gained from an assessment are related in large part to both the curriculum that is taught and the methods of instruction that are used (p. 51). In effect, curriculum, instruction, and assessment all have influence over each other. Pellegrino, Chudowsky, and Glaser define curriculum as "the knowledge and skills in subject areas that teachers teach and students learn" (p. 51). Instruction refers to the "methods of teaching and the learning activities used to help students master the content and objectives specified by the curriculum" (p. 52). Finally, assessment is defined by Pellegrino and colleagues as the "means used to measure the outcomes of education and the achievement of students with regard to important competencies" (p. 52). Assessment can happen in a variety of ways, including formal and informal means, classroom-based or large-scale, and formative (i.e., used to make decisions about instruction) and summative (i.e., used to measure outcomes).

Each level of the educational system—the classroom, school, district, state, or federal government—has different needs and uses assessment data in different ways and for different purposes. A classroom teacher may use the scores from her students' assessments in order to make decisions about instruction. At the school and district level, aggregated assessment scores may be used, for example, to target professional development and other

resources for teachers. This is one way in which schools may use assessment data for accountability purposes and to improve educational outcomes. At the state and federal levels, assessment data becomes increasingly about accountability.

Each level of education also plays a role in setting policy agendas and establishing policies for assessment, curriculum, and instruction, although the level of influence may be dependent on the type of decision being made, and who is making the decision. Pellegrino et al. (2001) have asserted that "states generally exert considerable influence over the curriculum, while classroom teachers have more latitude in instruction. States tend to determine policies on assessment for program evaluation, while teachers have greater control over assessment for learning" (p. 52). The federal government's role within this model is one of oversight and compliance. States have flexibility to establish their own policies and procedures, and standards and assessments, as long as the minimum requirements of the law are being met.

The Standards and Assessments Peer Review and the Policy Process

The No Child Left Behind Act is an educational policy that was passed by Congress in 2001. The standards and assessment peer review, begun in 2004, is an evaluation of states' implementation efforts. This process fits with

Dunn's (2008) description of the education policy cycle, illustrated in figure 2.2.

According to Dunn (2008), policy analysis is an intellectual activity "embedded in a social process" (p. 44) of making policy, which responds to a practical problem. The process of making policy is a cyclical, with each part of the process linked to the next. However, the process is not linear, and each part of the policy process can feed into another part or loop back, according to the needs and interests of stakeholders.

The No Child Left Behind Act places a strong emphasis on the policy outputs; that is, states should establish high quality standards and assessments directly aligned to those standards. States must create their own assessments, and also determine the ranges for proficiency on those assessments. However, it is the expectation of the federal government that by 2014, all students will be proficient on statewide assessments in English language arts, mathematics, and science. This is the policy output, according to Dunn's model.

The standards and assessment peer review is an evaluation of states' policy implementation efforts. States have established their own policies in order to develop their standards and assessments. They have, for example, worked to determine how students with disabilities and students who are

English language learners will participate in statewide assessments. States have made policy decisions about whether or not native-language assessments will be offered; states have established accommodations policies that support student achievement and maintain the validity of the assessment. The goal of the standards and assessments peer review is to determine whether or not the states' efforts comply with the expectations of the No Child Left Behind Act. If a state is found to be out of compliance with the Act, the state may make revisions to its policies, standards, and/or assessments in order to come into compliance with the Act. This is the process of adaptation, according to Dunn's model. For many states, this process is an ongoing, iterative process, as they struggle to meet the federal government's expectations for the standards and assessments with limited resources.

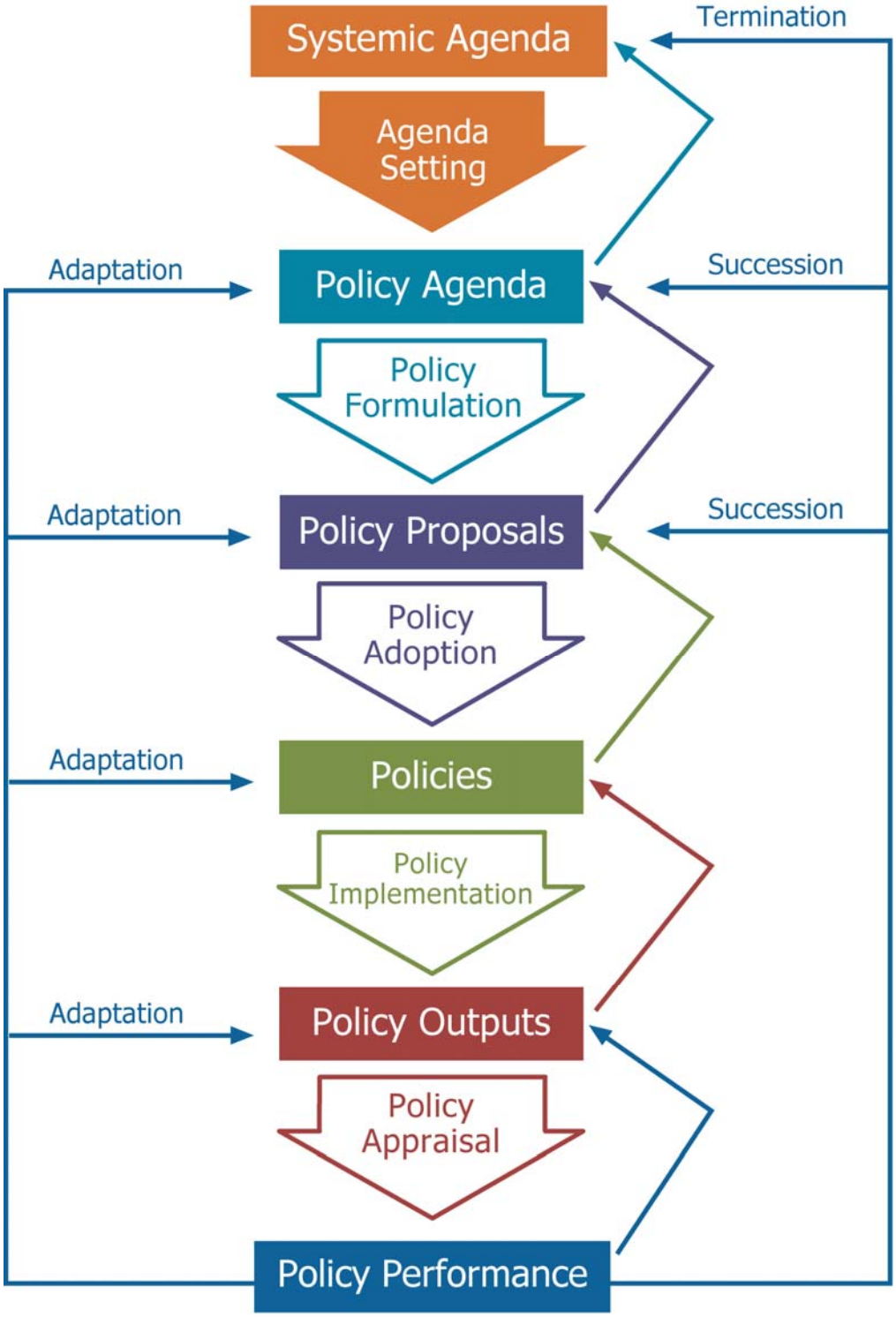

Figure 2.2 Dunn's (2008) illustration of the policy cycle

The No Child Left Behind Act of 2001 is a policy situated in an educational reform movement focused on increasing educational effectiveness. That is, while previous educational policies were focused on process (e.g., access to education for students from disadvantaged groups), NCLB focuses on outcomes. Creemers and Kyriakides (2008) suggest that educational effectiveness has many inputs: national policies, school-level policies, quality of teaching, and student-level factors. See Figure 2.3 for a detailed illustration of Creemers and Kyriakides' model of educational effectiveness, adapted to include ELLs by Liu (2009). It is worth noting that student-level factors includes a range of factors, including approach (e.g., aptitude, opportunity to learn), identity (e.g., gender, ethnicity), and attitude (e.g., expectations, and motivation). English language proficiency is included (adapted by Liu, forthcoming) under the attitude factors; however, this factor cuts across all three categories of student-level factors.

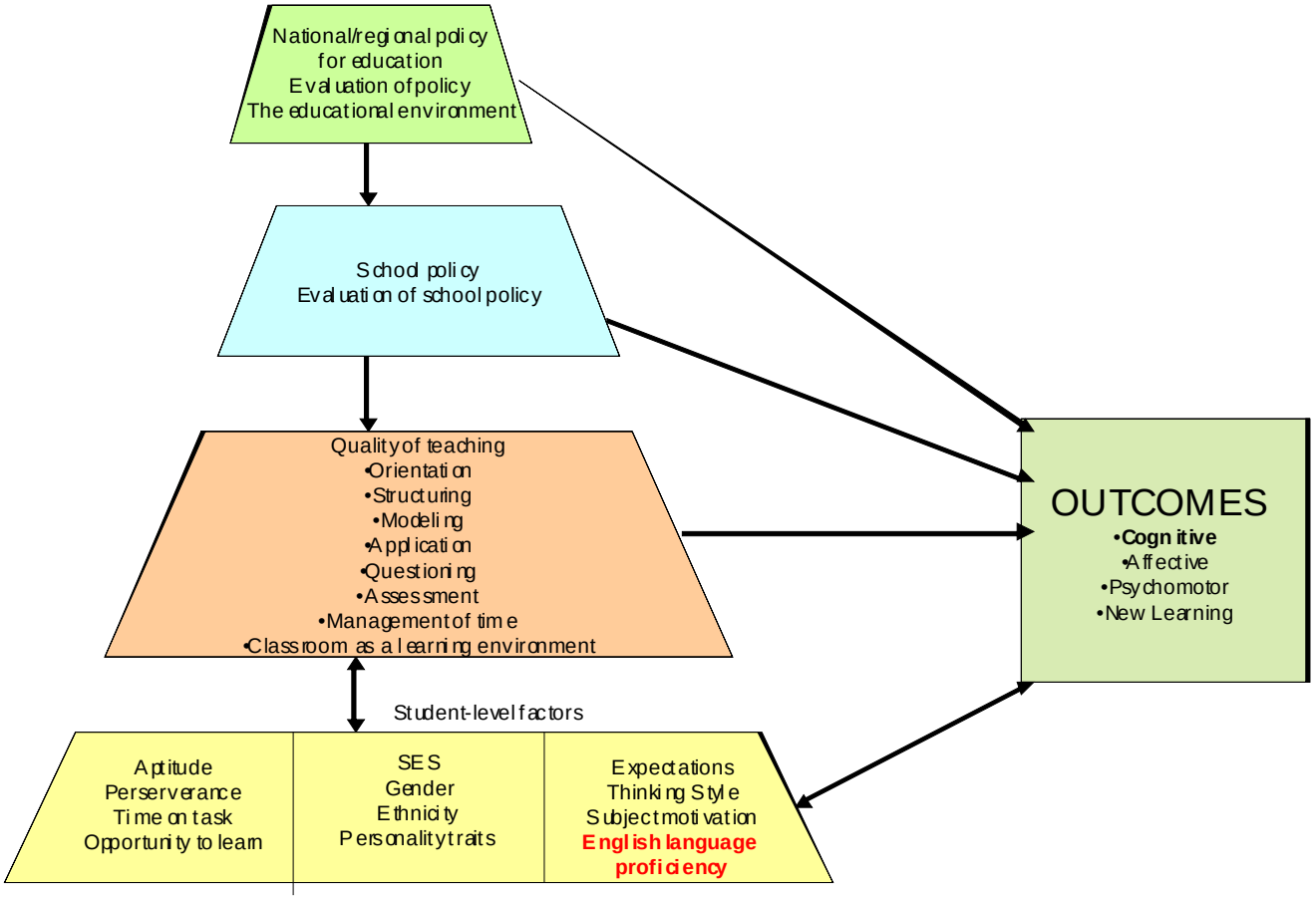

Figure 2.3. The dynamic model of educational effectiveness
(adapted from Creemers & Kyriakides 2008 by Liu 2009)

According to Creemers and Kyriakides (2008), national education policy such as NCLB is one factor leading to educational outcomes. However, national educational policy is a critical input, as it is the input from which all other inputs are established. The standards and assessments peer review is a method by which the federal government is able to check states' interpretation and implementation of this national policy. In this way, the federal government is able to ensure that the outcomes follow according to the policy as it was intended.

History of the Accountability Movement in U.S. Public Schools

Although No Child Left Behind was one of the first pieces of legislation under the George W. Bush administration and the law is often associated with his presidency, the standards-based accountability movement has been part of U.S. public education for many years, and large-scale achievement testing has been a central feature in education for even longer. According to Rigney, Wiley, and Kopriva (2008), large-scale testing in the United States began with the Army Alpha test, which was used to assess new army recruits. One of the authors of this test, Carl Brigham, was also credited with the development of multiple choice items. Soon, a version of this test was also being used for college admissions purposes (p. 38). By 1933, IBM machines were able to be used to score multiple choice tests, which made large-scale testing much

more cost-efficient, and as a result, this type of large-scale achievement test became the standard by the 1950s and continued to grow in usage through the 1980s (p. 38).

From the beginning, Title I of the Elementary and Secondary Education Act of 1965 (ESEA) has required "appropriate objective measures of educational achievement" (cited in Elmore & Rothman, 1999, p. 7). In response to ESEA, states began to increase the number of norm-referenced tests that students took. However, as Elmore & Rothman (1999) observed, norm-referenced tests do not provide information about what students know and can do. As a result, norm-referenced tests may be useful in making comparisons across states, but they do not help guide instruction.

In 1989, the National Council of Teachers of Mathematics published standards for mathematics education (Elmore & Rothman, 1999, p. 15), and five years later, the 1994 reauthorization of the Elementary and Secondary Education Act reflected this increasing value on standards-based education reform. The 1994 Title I law required states to develop standards and assessments, and that these standards and assessments needed to be the same for all students, whether or not they were eligible for Title I (Elmore & Rothman, p. 8). There have been strong consequences to this shift toward standards. Pellegrino, Chudowsky, and Glaser (2001) have noted that the

standards movement has “increased the amount of testing in K-12 schools and raised the consequences, expectations, and controversies attached to the test results” (p. 24). However, unlike previous emphases on testing, the new law required student test scores to be compared to the standard, rather than to other students. Under this model of assessment, when assessments are constructed to measure what a student knows and can do, all students can demonstrate proficiency when compared to the standard.

The primary change from the 1994 reauthorization of ESEA to the 2001 reauthorization (NCLB) is that under the new law, states are held accountable for developing standards and assessments, and ensuring that all students are scoring at or above proficient in reading, math, and science. ESEA is currently up for reauthorization, although it is likely that no changes will occur until 2009 at the earliest. Nonetheless, education policymakers do not predict major changes in the law as it is today.

One way in which states must demonstrate that they have challenging content standards, appropriate achievement standards, and assessments is to have their standards and assessments go through a federal peer review process. The standards and assessment peer review process, described briefly in Chapter 1, has been the subject of some limited research. Christensen et al. (2007) and Thurlow et al. (2008) examined the peer review comments in

relation to issues of importance to accommodations for students with disabilities. Bechard (2008) considered the peer review decision letters in order to help states better understand the key issues for developing alternate assessments that would pass the peer review. Finally, the General Accounting Office (2007) reviewed the peer comments in order to provide testimony to Congress regarding assessment issues for English language learners.

This study by the GAO is different from the current study in several ways. The study by the GAO was conducted in order to provide testimony to Congress rather than to provide an in-depth analysis of issues emerging from the peer review comments. The GAO study considered multiple data sources in addition to the peer review, such as state policies and reported assessment data. Finally, another difference is that the GAO study was conducted at an earlier point in the peer review, and thus does not include data that is as comprehensive as the data used for the current study.

As mentioned previously, the standards and assessments peer review has required states to ensure that the followed items have been addressed in each state's accountability system:

(1) A single statewide system of challenging academic content standards applied to all public schools and LEAs;

(2) A single statewide system of challenging academic achievement

standards applied to all public schools and local education agencies;

(3) A single statewide system of annual high-quality assessments;

(4) A system of assessments with high technical quality;

(5) Alignment of academic content standards, academic achievement standards, and assessments;

(6) Inclusion of all students in the assessment system; and

(7) An effective system of assessment reports (U.S. Department of Education, 2004, pp. i-ii).

In each of these areas, there are key components relevant to the assessment of English language learners. In the following pages, each of these sections will be defined in more detail and the relevance of each section to the assessment of English language learners will be explored along with a review of literature relevant to each topic.

Section 1: A Single Statewide System of Challenging Academic Content Standards

With regard to Section One which relates to the development of content standards, the US Department of Education Office of Elementary and Secondary Education (OESE) (2004) stated that challenging academic content standards are the basis for a high quality assessment and accountability

system. These standards are intended to “define what all public school students in [each] State are expected to know and be able to do” (p. 8).

English language learners are affected in two ways by this Section: first, the content standards developed by states must apply to *all* students, including English language learners; and second, the development of state content standards must include involvement by relevant stakeholders, including parents and others familiar with students who have limited English proficiency.

Including all students in the accountability system has proven challenging over time. Researchers who focus on testing ELLs (see for example, Abedi, 2007; Kopriva 2008; Liu and Anderson, 2008) noted that English language learners are a very diverse group—they vary by home language and cultural background, socio-economic status, years of schooling, and disability status, among other variables. Furthermore, the population of English language learners is increasing. As Abedi (2002) observed, the number of English language learners in schools in Texas, California, Florida, and New Mexico has exceeded 10% of the school population. This may be expected, as these are states along the border with Mexico. However, other states have seen a marked increase in the population of English language learners in schools as well. For example, the population of students who are

considered LEP in Minnesota has risen from 24,962 in the 1995-96 school year, which was almost 3 percent of the school population, to 59,127 in the 2005-06 school year, which was roughly 7 percent of the school population. This represents 136.9% growth over the last ten years (NCELA, 2008). Other states have seen similar patterns of growth, and this increase in English language learners has implications for the development of student assessments.

Cummins (2000) has expressed concern that in developing standards and assessment systems, LEP students have been included more as an afterthought. For example, he noted that when determining the validity of an assessment, English language learners have not always been taken into consideration (p. 146). Cummins reported that some supporters of LEP students have therefore been proponents of excluding them from accountability systems because the assessments may not be equitable and may not yield valid results for LEP students. Similarly, test developers and schools have found it easier to exclude English language learners and other groups from the assessment rather than to develop tests and testing conditions that are inclusive. However, Cummins reported that participants in the National Clearinghouse for Bilingual Education conference ultimately acknowledged that LEP students should be included in accountability testing

as long as they are provided appropriate accommodations and alternative means of testing so that the results can be both meaningful and useful (p. 148). Furthermore, Cummins strongly advocated that there must be an ongoing research agenda relevant to the assessment of English language learners.

Although there has been some resistance among educators to including LEP students and other disadvantaged groups of students in accountability systems, Thurlow and Ysseldyke (2002) reported that there are several reasons why educators should want to include all students in accountability systems. These reasons include the following:

(1) A more accurate picture of how students are doing in the educational system is obtained when all students are included;

(2) Fairer and more accurate comparisons are possible only if all students are included in comparisons;

(3) Inclusion of all students meets the letter and intent of several federal laws; and

(4) Policy decisions consider all students rather than a subset of students (p. 20).

Furthermore, students themselves benefit from inclusion in accountability assessments. When students are included, they are more likely to receive

instruction on standards-based content, they are able to learn test-taking skills, and they are held to higher expectations (Thurlow and Ysseldyke, 2002, p. 20).

No Child Left Behind supports this position of Thurlow and Ysseldyke. In Section 1 of the peer review, states must demonstrate that they have a single statewide system of challenging content standards, and that these standards apply to all students. No students may be excluded from the accountability system according to the provisions of No Child Left Behind, and states must demonstrate that "all means all."

Section 2: A Single Statewide System of Challenging Academic Achievement Standards

With regard to this section, the peer review guidance document stated the following:

> To establish the level of achievement a State expects of all public schools and LEAs, the NCLB requires States to develop a set of challenging academic achievement standards for every grade and content area assessed. These standards are to be applied to all public schools and LEAs and ensure inclusion of those students with disabilities and students who are not yet proficient in English (U.S. Department of Education, 2004, p. 14).

The Department of Elementary and Secondary Education continued in its description of this section, adding that state documents pertaining to achievement standards must "specifically state that its academic achievement standards apply to all public and elementary and secondary students, including students with disabilities and students who are not proficient in English" (p.20). Similarly, OESE observes that the word "all" must be used by states in their documents to demonstrate the inclusion of *all* public school students. Additionally, this Section also requires a diverse panel of participants, including parents and others knowledgeable about students who are English language learners, in the development of the achievement standards. Above all, this section requires that schools attend to the achievement of all students; this attention to the achievement of ELLs has been a contentious political issue that began through civil rights legislation and has continued through to No Child Left Behind.

ELLs and the Achievement Gap

There is a well-documented achievement gap between white students in the United States and their language minority counterparts (e.g., Cummins, 1984; Cummins, 1989; Durán, 1989; Ortiz &Wilkinson, 1991; La Celle-Peterson and Rivera, 1994; Lazarín, 2006; Fry, 2008; Batalova, Fix, and Murray, 2007; Fry, 2007 Wilkinson & Ortiz, 1986). For example, Lazarín

(2006) cited data from the 2005 National Assessment of Educational Progress (NAEP), noting that

> 29% of eighth-grade ELLs scored at or above the basic achievement level in reading, compared to 75% of non-ELLs. Similar gaps were demonstrated in mathematics. These gaps are also evident at the state level. In Texas, in the 2001-2002 school year, ELLs in grades 7-12 were retained in grade at twice the rate of English-proficient students (13% compared to 6.5%, respectively) and dropped out of school at a rate of 77% greater than non-ELLs. (p. 4)

This is one example of the achievement gap between ELL students and their native English counterparts.

Other researchers have considered the achievement gap with regard to the content assessments required of NCLB. For example, Fry (2008) analyzed the publicly reported data from states' reading and mathematics assessments, and reported the following results:

> In both reading and mathematics, a majority of ELL students who took assessment tests scored below proficiency standards. Furthermore, ELL test-takers were far behind the other major racial/ethnic groups in measured achievement in elementary school. The gaps with other major racial/ethnic groups widen from elementary grades to eighth

grade. (p. 3)

The achievement gap has been characterized as a civil rights issue, and has been addressed over the last fifty years by both legislation and legal decisions.

ELLs and Civil Rights

Although conflict over language politics in the United States can be traced at least to colonial times (Crawford, 2000), concern about language issues in education for English language learners really begins with civil rights legislation of the twentieth century and through some court decisions. The Civil Rights Act of 1964 (42 USC §2000d, cited in Crawford, 2000) stated that "no person can be excluded from participation in or denied the benefits of" federally supported programs on the basis of race, color, or national origin. The implications of this legislation were slow to be interpreted in the education context (Crawford, 2000), but provided a foundation of civil rights for ELLs.

The Elementary and Secondary Education Act of 1965 itself was seen to be part of a movement of civil rights legislation in relation to education, as it increased opportunities for students who were considered to be minorities. However, in its original form, ESEA did not include ELLs. In 1968, the Bilingual Education Act was passed in order to provide funding to schools to increase their resources for language minority students, with the intention of closing

the achievement gap for LEP students. The Bilingual Education Act has also been known as Title VII of ESEA.

Although these pieces of legislation were designed to provide resources to schools and districts to close the achievement gap for LEP students, school districts did not necessarily implement policies in relation to any of these civil rights changes. One noticeable example was the San Francisco school district, which provided identical English instruction to Chinese-speaking students. In the landmark Supreme Court suit brought by the parents of one such student, *Lau v. Nichols* (1974, cited in Crawford, 2000) became a key ruling in support of English language learners, and, as Crawford asserted, this ruling forced school districts to pay attention to the needs of LEP students, as the court found that identical instruction is not equal instruction.

In response to the *Lau v. Nichols* case, the Bilingual Education Act was revised several times throughout the 1970s and 1980s in order to define bilingual education, as well as to establish the means by which LEP students would be served by the Act. In the 1980s, the Reagan and Bush administrations limited funding for bilingual education, and reestablished bilingual education as programming for English as a second language.

In 1981, a federal appeals court ruled on another case in support of LEP students. *Castañeda v. Pickard* gave schools a three-pronged approach in

order to gauge the effectiveness of their efforts to remove language barriers:

(1) Programs must be based on an educational theory recognized as sound by experts.

(2) Resources, personnel, and practices must be reasonably calculated to implement the program effectively.

(3) Programs must be evaluated and restructured, if necessary, to ensure that language barriers are being overcome.

This ruling provides one means of holding schools accountable for the achievement of ELLs by attending to both inputs and outputs. These three prongs are echoed in NCLB.

In the 1990s, as the standards movement that would lead to No Child Left Behind was being implemented, there was a considerable backlash against bilingual education programs, which had already lost much federal funding under the Reagan and Bush, Sr. administrations. The “English Only” movement rose in popularity (Crawford, 2000), and legislation, such as Proposition 220 in California which passed in 1998, and Proposition 203 in Arizona which passed in 2000, have required that English be the language of instruction in these states. These propositions have effectively eliminated bilingual education programs in these states, and other states have passed similar laws (Crawford, 2007). Although these laws have frustrated and

angered many supporters of English language learners, some proponents of these laws intend that English Only programs are a means to address the achievement gap by giving LEP students a short time to focus on English language instruction before moving to direct content instruction.

NCLB itself is an attempt to close the achievement gap for LEP students as well as other disadvantaged students. In this particular section, by requiring all students to be held to the same achievement standards, NCLB intends that schools will ensure that all students are given access to high quality instruction. It should be noted that in this section of this literature review, civil rights legislation and relevant court cases have been discussed as they pertain to closing the achievement gap for LEP students. However, the peer review process itself does not address whether or not states are in compliance with civil rights legislation.

Rather, the peer review, as it relates to this section, examines only that states have high quality academic achievement standards and that all students are held to these standards. Ensuring that the achievement standards are established with input from stakeholders that include those with the interests of LEP students in mind, for example, is one step toward making sure that the achievement standards are fair and equitable to LEP students.

Section 3: A Single Statewide System of Annual High-Quality Assessments

Section Three required states to have a comprehensive assessment system and mandated that all students must participate in this statewide assessment system. In addition, this section described the federal requirements for testing: that states must have assessments for English language arts and mathematics in place by the school year 2005-06, that students be tested in grades 3-8 and once in grades 10-12, that science assessments must be included in the assessment system by school year 2008-09, and that assessments must be ones that ***all*** (OESE emphasis) student must take (p. 23). Thus, a key issue for English language learners with regard to this section is how students who are designated as LEP will participate in statewide assessments.

Importance of Participation

The participation of all students in assessments is a cornerstone to the accountability provisions of NCLB. By including all students in the assessment system, states are ensuring that students have access to the high quality instruction they need in order to perform their best on the assessments. For example, prior to NCLB, ELLs were regularly exempted from statewide testing, which lead to concerns about how much access ELLs had to high quality instruction. In Massachusetts, "78% of ELL sophomores were exempted from the state's reading/English language arts test in 1999-2000. Student

performance data about these students, as a result, were limited as well, and the little data that were collected revealed dismal academic outcomes" (Albus, Thurlow, and Liu, 2002, reported by Lazarin, 2006, p. 7). In another example, only 2% of the ELL students in Florida who were assessed in English language arts scored as proficient in the school year 1999-2000 (Albus, Thurlow, and Liu, 2002). Given these poor outcomes for English language learners, NCLB's participation requirement has been seen as a remedy to this situation.

Ways of Participating

Students have several ways in which they may participate: most students take the statewide general assessment without accommodations; some students with disabilities and English language learners may take the general assessment with accommodations; in some states an ELL may be able to take an alternative version of an assessment, which usually indicates a native language assessment; and a very small number of students may take an alternate assessment. States must have an alternate assessment available for the very small number of students with significant cognitive disabilities who are unable to participate meaningfully in the general assessment even with accommodations. These various participation options will be discussed in more detail under Section 6: Inclusion.

Participation and Identification of English Language Learners

In addressing the inclusion of English language learners with regard to this section, states must demonstrate that English language learners are included as participants in the assessment system. Researchers (2008) have pointed out that one challenge states have with regard to the participation of ELLs in statewide assessments is with how students are identified as ELLs. Wolf et al. (2008) observed that identification procedures vary widely from state to state. Durán (2008) noted that many states use a family questionnaire to determine the language background of a student, with most states using a follow-up English language proficiency assessment if a parent notes that a language other than English is spoken in the home. Durán stated that some school districts may be motivated to have greater numbers of ELLs in their schools just as other districts may see a benefit in being more conservative in placing students into ELL/LEP designations. However, no research has been conducted on the placement process itself (Durán, 2008, p.302), which is unfortunate given that a student's designation as ELL/LEP has implications for which type of assessment (general, alternative, or alternate) a student may take as well as whether or not a student may be eligible to receive accommodations, depending on the state.

Under NCLB, a student who is designated as LEP may be exempted

from accountability testing in English language arts for one year. However, LEP students may not be exempted from mathematics assessments. An additional challenge as it relates to the identification of LEP students is that some students in this subgroup may also be migrant students, and may not stay in the same school district for an entire academic year. This may also be the case for other LEP students who are not specifically designated as migrant students. For schools, there is a challenge in knowing when a student can and cannot be exempted from English language arts testing.

Section 4: A System of Assessments with High Technical Quality

Section Four refers to the technical quality of the assessments. Technical quality, according to the *Standards for Educational and Psychological Testing* (1999) is the process used in the development of an assessment to ensure that results are "appropriate, credible, and technically defensible" (cited in OESE, p. 32). Technical quality includes the assurance that test scores are valid and reliable, that scores provide information on what a student knows and is able to do, that the questions are fair and free from bias, and that the test is accessible to all students. Thus, several components of technical quality relate to English language learners. States must make sure that linguistic accommodations are available to LEP students, that assessments have been developed to be fair and free from bias, and that

accommodations use yields valid and meaningful scores. States must demonstrate that they have evaluated the use of linguistic accommodations so that the use of accommodations for LEP students can yield reliable information about what students know and are able to do. In addition, states must demonstrate that assessment accommodations result in scores that are comparable to test results from non-accommodated administrations. This section, then, addresses three major issues with regard to ELLS: 1) accommodations for ELLs, 2) bias reviews, and 3) analyses of differential item functioning (DIF).

Accommodations for ELLs

Assessment accommodations are changes in the materials and procedures of an assessment that allow an English language learner to demonstrate what she knows and can do in relation to the content of the assessment. Assessment accommodations allow "the student's true knowledge and skills to be assessed" (Thurlow, Lazarus, and Christensen, 2008, p. 18) rather than the student's proficiency in English. Ultimately, accommodations improve the validity of interpretations that can be made about the student's score when accommodations are used by the students who need them (Sireci and Pitoniak, 2007, p. 53).

Accommodations may be used by students with disabilities as well as

English language learners. In terms of accommodations for ELLs, Rivera, Collum, Willner, and Shia (2006) developed a taxonomy in order to classify accommodations as either direct linguistic supports, such as a bilingual dictionary, or indirect linguistic supports, such as extended time. This taxonomy has been particularly important in order to differentiate ELL accommodations from those intended for students with disabilities, as in most cases the policies for accommodations for students with disabilities were developed prior to policies for accommodating ELLs. Given that there has already been shown to be a tendency to over identify ELLs as special education students, this distinction has been important in making sure that ELLs have full access to the general curriculum (Rivera and Collum, 2006). When accommodations are seen as helping students overcome language barriers, they have been determined to be "ELL-responsive" as defined by Willner, Rivera, and Acosta (2008). Of the 104 accommodations Willner et al. identified in state policies, only 40 of these were found to be "ELL-responsive" (p. 14). Commonly used "ELL-responsive" accommodations include direct translations of the test directions and/or questions, use of bilingual dictionaries and glossaries, and use of word lists.

Research on Accommodations for ELLs

Given the potential number of ELL accommodations available to students, there is a general paucity of research on the effectiveness of accommodations for ELLs. A meta analysis conducted by Pennock-Roman and Rivera (2007) revealed a total number of 8 experimental studies on ELL accommodations. These accommodations include the following: (1) English glossary; (2) English dictionary; (3) plain English, or simplified English, version of test; (4) dual language glossary; (5) dual language word-to-word dictionaries; (6) provide side-by-side dual language tests; (7) translated tests, and (8) extended time (Pennock-Roman and Rivera, 2007). The majority of ELL accommodations actually used in states have not been researched (Willner et al., 2008). This is important to states, as the peer review guidance requires that states demonstrate that scores from accommodated assessments will yield valid results that can be compared to scores from un-accommodated administrations. Research on the effectiveness of these accommodations is a valuable component to assuring validity.

Policies on Accommodations for ELLs

Establishing sound policies for assessment accommodations is another means by which states can maintain validity for accommodated assessments. In their review of state accommodations policies, Willner et al. (2008) found

that all 50 states and the District of Columbia offered at least one accommodation for English language learners. However, they found that a range of possibilities for accommodations existed, with at least one state offering as few as 2 accommodations for ELLs and other states offering as many as ELL-responsive accommodations (p. 20).

While all states have some established policies for ELL accommodations, not all states include guidance for decision-makers to assign accommodations to students. Willner et al. (2008) found that only 33 states provided some guidance for decision-makers (p. 27). This is in contrast to policies for students with disabilities, in which all 50 states provide some policy guidance for teachers and other decision-makers to assign accommodations to students with disabilities (Christensen, Lazarus, Crone, and Thurlow, 2008).

Bias Review

Another means by which a state can show that they are attending to the validity of assessments for diverse students is to conduct a bias review of test items. Hambleton and Rodgers (1997) define bias as "the presence of some characteristic of an item that results in differential performance for individuals of the same ability but from different ethnic, sex, cultural, or religious groups" (p.1). In conducting a bias review, states intend to make sure that all test items are fair, and that they are free from bias and

stereotyping. From the standpoint of NCLB, the Department of Education considers bias review to be an important element in maintaining the validity of the assessment because when questions are free from bias, students are able to approach the questions equitably.

States regularly conduct bias reviews of their assessment questions. The results of these reviews are generally not made public; however, the participations in bias reviews is an important aspect of the process itself. When minority students are appropriately represented in the bias review process, potential areas of bias are more likely to be uncovered.

Differential Item Functioning (DIF)

After a bias review has been conducted, another means in which state test developers can ensure that test items are fair to all students is to conduct a differential item functioning (DIF) analysis. Zumbro (1999) explained DIF analysis as follows:

> DIF statistical techniques are based on the principle that if different groups of test-takers (e.g., males and females) have roughly the same level of something (e.g., knowledge), then they should perform similarly on individual test items regardless of group membership. In their essence, all DIF techniques match test takers from different groups according to their total test scores and then investigate how the

> different groups performed on individual test items to determine whether the test items are creating problems for a particular group. (p. 1)

Through this process of pilot testing, actual items are given to students from different demographic groups. If the students' performance is seen to be different based on a systematic source of error such as their identification with a particular group, then that test item is seen to be problematic.

Similar to providing accommodations, DIF analysis is conducted in order to ensure that the assessment is testing what it intends to, rather than some other unintended construct. In addition, DIF analysis helps test developers and other stakeholders know that the test is fair and equitable to all students who participate in the assessment.

States and test vendors regularly conduct DIF analysis; however, the results of these analyses are rarely made publicly available. This is largely because state agency personnel are unable to prioritize making these studies publicly available through peer journals or other means. Thus, the results of these studies as they pertain to ELLs are not discussed here. However, for peer review purposes, states have been able to submit these technical reports as evidence for the technical quality of their assessments.

Section 5: Alignment of Standards and Assessments

Section Five focuses on the alignment of achievement standards with assessments. Alignment ensures that assessments measure what they are supposed to measure: a student's knowledge based on the state's content standards. If assessments are not aligned to standards, the assessment may measure something other than a student's knowledge of the content or the assessment may do a poor job of measuring the student's knowledge. Although alignment of standards and assessments has implications for ELLs, these implications relate more directly to their inclusion in the educational system as well as their participation in the assessments. These topics were discussed in more detail in sections one, two, and three, and therefore, will not be re-examined here.

As an addendum to this section, the peer review guidance document added a note about informing stakeholders of the importance of alignment:

> the State needs to also develop strategies for communicating to its education stakeholders how its standards and assessment are aligned. Parents, educators, and other stakeholders need to know how assessment results are related to content-based expectations in order to understand and use test information effectively (U.S. Department of Education, 2004, p. 41).

This issue of communicating effectively with stakeholders, including parents, is covered by the peer review guidance in more detail in section 7. Thus, this topic is also explored in more detail here as it pertains to section 7.

Section 6: Inclusion of All Students in the Assessment System

Section Six relates to the inclusion of all students in the assessment system. In this section, OESE returns to one of the major premises of NCLB—high expectations for all children:

> Just as its title indicates, one of the fundamental principles of the NCLB is the inclusion of ***all*** students in a state's system of standards, assessments, and accountability. By excluding any student or group of students from its assessment system, a state suggests that its high expectations apply only to some, but not all, students (p. 50).

As was also explained under Section 3 of this literature review, states must have participation options available for LEP students, including participation in the general assessment, participation in the general assessment with accommodations, and states must offer at least one alternate assessment. In addition to an alternate assessment, states may also have different forms of an assessment, such as a native language or simplified English assessment. These are often referred to as alternative assessments. In this section, alternate assessments will be discussed, followed by alternative

assessments for ELLs.

Alternate Assessments

In addition to taking the general assessment, with or without accommodations, another means by which an ELL student may participate in statewide assessments is by taking an alternate assessment. An alternate assessment may be an alternate assessment based on alternate achievement standards. That is, this assessment would have different achievement standards for the students who take them. These assessments are intended to be used for students with the most significant cognitive disabilities. The federal government allows states to include as proficient one percent of all students tested at each grade level in this alternate assessment category.

While most states only offer an alternate assessment based on alternate achievement standards for students with the most significant cognitive disabilities, states also have the option to develop an alternate assessment based on grade level achievement standards; however, this assessment is currently only used in two states. A third option for an alternate assessment is an alternate assessment based on modified academic achievement standards. This type of alternate assessment is designed for students with disabilities who may not have a significant cognitive disability but for whom the general assessment may be deemed inappropriate. An

additional two percent of all students tested at each grade level may be counted as proficient according to this type of alternate assessment, which is based on modified academic achievement standards. States are currently in the process of developing this type of assessment, and many states see this alternate assessment as a viable option for ELLs who also are classified as having a disability.

Zehler, Fleischman, Hopstock, Pendzick, and Stephenson (2003) found that roughly 9% of the ELL population is identified as having a disability. Most ELLs with disabilities are classified as having a speech or language impairment, although ELLs with disabilities are represented in all thirteen categories of disability. Similarly, most ELLs with disabilities speak Spanish as a native language, according to Zehler et al. (2003). However, many language backgrounds are represented among ELLs with disabilities.

English language programs and special education programs have had a complex relationship. On the one hand, LaCelle-Peterson and Rivera (1994) have pointed out that English language learners have been often over-referred to special education programs. This has often been the case in schools that had limited supports for English language learners. On the other hand, Shyyan, Thurlow, and Liu (2008) noted that English language learners who do have disabilities often take longer to be referred to special education

programs. Often, this may be due to educators wanting to ensure that a particular student's challenges are not exclusively one of language learning. However, in some cases, the decision to refer or not to refer an LEP student to special education may be political.

Anecdotal evidence has suggested that the number of LEP students, with or without disabilities, taking alternate assessments is likely to be disproportionately high in comparison to the number of native-English peers taking alternate assessments. Unfortunately, however, Albus and Thurlow (2005) have observed that most states do not report participation or performance data for ELLs with disabilities on their content assessments. The extent of this issue is currently unknown.

Alternative Assessments

In addition to alternate assessments, states may also provide alternative assessments. These assessments are a different format of the same general assessment. They may be a native language assessment, in which the assessment has been translated into a language other than English. Other states may use an assessment in simplified English; this assessment type removes difficult vocabulary and sentence structures that may distract the student from the actual content being assessed.

Native Language Assessments

In some states, educators have developed assessments in languages other than English as one means to attend to both participation rates of ELLs as well as concerns about an ELL's ability to demonstrate his or her achievement in English. A native language assessment is one type of alternative assessment; that is, it is measuring the same content constructs as the English-language assessment, but the assessment is in a different language. This is usually done in one of two ways. Either the test is provided in a direct word-for-word translation into the native language, or the translated test is provided as a side-by-side version with an English language assessment (Bolt and Roach, 2009).

Several challenges have been identified with regard to native-language assessments. The development of an assessment is very expensive, and a native-language assessment is considered to be an entirely separate assessment, rather than simply a translation of an English-language test. In addition, the question of equity may be raised with regard to what language the assessment may be written in. Most of these alternative assessments are in Spanish due to the relative size of the population of students who may be eligible to take such an assessment. However, other students who speak a less commonly spoken language may also benefit from an assessment in their

native language. Finally, research has found that assessing a student in a language other than English is only appropriate when the student receives instruction in that language (Shyyan et al., 2008; Abedi, 2001).

Simplified English Assessments

Another approach that states may take so that LEP students can demonstrate what they know and can do on statewide assessments is to use a simplified English assessment. Rivera and Stansfield (2001) reported that simplified English grew out of a business model developed by the Caterpillar Tractor corporation in the 1970s and 1980s. Simplified English often involves using limited core vocabulary alongside a glossary that contains necessary technical terms. The structure of the language is attended to, so that only active verb forms may be used, for example.

Only a few studies have been conducted on the use of simplified English for statewide assessments (Abedi, 1997; Kiplinger, Haug, and Abedi, 2000; Albus, Bielinski, Thurlow, and Liu, 2001; and Rivera and Stansfield, 2001); however, these studies tend to support the use of simplified English assessments. For example, Abedi (1997) found that LEP students who were native Spanish speakers performed better on a simplified English assessment than those native Spanish speaking LEP students who took a Spanish version of the assessment.

Considerations for Using Alternate and Alternative Assessments

In addition, states must show that scores on each assessment are comparable across subgroups, including English language learners. Furthermore, if states provide native-language or simplified English versions of an assessment, they must provide evidence to prove that the scores on these versions of the assessment are comparable to other versions. These alternative versions must also demonstrate that they are aligned to the state's content and achievement standards as well.

Section 7: An Effective System of Assessment Reports

Section Seven addresses reporting assessment results. The peer review guidance stated the following about reporting:

> A State's assessment reports represent the culmination of all other aspects of its standards and assessment system. In these reports, a parent, educator, or other stakeholder should find answers to questions about how well a student or group of students is achieving, as well as important information on how to improve achievement in the future. (U.S. Department of Education, 2004, p. 58)

When reporting assessment results to stakeholders, several considerations must be followed. States must provide reports at a number of levels: individual student, school, district, and state. Assessment scores must be

aligned to content standards, and scores must be reported relative to achievement levels. In addition, states must provide interpretive guidance so that parents and other stakeholders less familiar with assessments may understand the information contained in these assessment reports. Interpretive guidance must be made accessible to all stakeholders (U.S. Department of Education 2004, p. 58).

This section has two major elements relevant to English language learners. While assessment results must be publicly reported, states must also ensure that when subgroups are small enough to reveal a student's identity, these subgroups are not reported. This element does not pertain only to English language learners, but it can be relevant to them in schools where there are only a few ELLs in each grade. More directly relevant to ELLs is the requirement that individual student reports must be displayed in a format and language that parents can understand. In addition, parents must be given interpretive guidance.

Minimum "N-Size" and Accuracy of Reporting

With regard to reporting scores at all levels except the individual, care must be taken so as to not violate the confidentiality of the student. Thus, in schools in which the population of any subgroup, such as ELLs, is small enough that individual students could be identified, schools and districts do

not have to report the results for that subgroup (U.S. Department of Education, 2004, p.58). This is commonly referred to as the "minimum n-size" which refers to the number of students in the population of a subgroup for reporting purposes. There has been wide variability with regard to the minimum number of students established by states for reporting purposes. Some states established policies that the minimum "n size" could be large for some subgroups and smaller for other subgroups.

Lazarín (2006) stated that one problem associated with the minimum "n size" is that states could "game the system" in order to "mask the low achievement of Latino and ELL students by reporting aggregate achievement results. Furthermore, ELLs and students with disabilities were often excluded from testing completely, and when they were included in testing, their scores often went unreported. For example, according to data from the 1998-1999 school year, ELLs were able to be exempted from statewide testing in the policies of 46 states (Rivera, Stansfield, Scialdone, and Sharkey, 2000, reported in Lazarín, 2006). Under the current provisions of NCLB, ELLs may only be exempted from English language arts assessments in their first year of schooling.

At the time that the peer review process was initiated, there was no specific requirement regarding the minimum "n-size." However, in April 2007,

new regulations established that states are required to set one consistent "n-size" for all subgroups reported (U.S. Department of Education, 2007).

Accessible Language

Reporting the results of assessments in language that is accessible to all stakeholders—especially parents—is a cornerstone to the accountability provisions of NCLB. As Lazarín (2006) explained:

> By arming parents and the public with information about their neighborhood schools and state education system, NCLB seeks to increase the likelihood that parents will engage in, monitor, and hold their child's school accountable for classroom learning and outcomes. (p. 7)

NCLB requires states to make information on assessment scores accessible to parents and other stakeholders, and states are able to make their own determinations about how to best accomplish this goal.

Reporting assessment scores is particularly relevant to ELLs because their parents may not be fluent in English. In order to make this information accessible to parents, Lazarín noted that some states report assessment results in English and in Spanish (p. 16). However, most states post assessment scores on their department of education websites, which must be navigated in English and also require access to the Internet.

Conclusion

With the exception of section 5, each of the sections of peer review guidance has components that directly relate to the assessment of English language learners, and section 5 has indirect relevance to ELLs. With so many issues to contend with, it is not surprising that as of December 2006, only eighteen states had received approval of their standards and assessment systems. A comprehensive review of the issues from the peer review that are important to assessing ELLs is needed. Such a review may indicate what assessment issues are most prevalent in the comments, and therefore which issues should take priority. Furthermore, for states that still have outstanding issues in the peer review, such a comprehensive review may help state agency personnel know how to better address these issues as they subsequently continue to submit evidence the review process. Finally, all of this information can be used to improve the overall assessment of ELLs and to ultimately improve student learning.

Thus, the research questions this study proposes to answer include the following:

1. What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be

in compliance with the standards and assessments accountability requirements of No Child Left Behind?

2. What are the commonalities among states in meeting the federal requirements for assessing English language learners?

3. What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest?

The following chapter describes the methodology that will be used to answer these three questions.

CHAPTER III: METHODOLOGY

Research is formalized curiosity. It is poking and prodding with a purpose.
--Zora Neale Hurston

The purpose of this chapter is to describe the methodology and research methods used in this study. In the first part of the chapter, the research methodology used for the study and the rationale for its use are described. The second part of the chapter highlights the procedures used to collect the data. The third part describes methods of data analysis. Finally, the concluding part focuses on the limitations of the research methods used.

This research study is a secondary analysis of the federal process to monitor the policy outcomes related to the standards and assessments portions of NCLB. The peer review of standards and assessments falls under "compliance monitoring," defined by Dunn (2008) as monitoring to "determine whether the actions of program administrators, staff, and other stakeholders are in compliance with standards and procedures imposed by legislatures, regulatory agencies, and professional bodies" (p. 274). The federal review process was the actual determination of whether or not a state's standards and assessment system are in compliance with NCLB. This research study proposes to focus attention on the issues relevant for ELLs that emerged from this compliance monitoring procedure.

In the case of the Standards and Assessments Peer Review, the compliance monitoring is focused on process, rather than outcomes. That is, the intent of the review is not to evaluate the outcomes of states having standards and assessments, but instead that states have completed all the necessary steps to ensure that they have appropriate standards and assessments in place. In this regard, the Standards and Assessments Peer Review is monitoring the implementation of educational policy.

Mixed Methods Approach

The mixed methods approach is one in which both qualitative and quantitative frameworks are used to collect and analyze data (Creswell & Plano Clark, 2007; Tashakkori &Teddlie, 2003. Mixed methods can be a preferable approach when the research questions cannot be answered using either qualitative or quantitative methods alone. In the case of the study undertaken here, a mixed methods approach is required, as the research questions require both qualitative and quantitative methods be employed in obtaining the answers.

Research Questions

The research questions to be answered by the current study include the following:

1. What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be in compliance with the standards and assessments accountability requirements of No Child Left Behind?

2. What are the commonalities among states in meeting the federal requirements for assessing English language learners?

3. What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest?

The first question requires the use of quantitative methods. To answer this question, a descriptive quantitative analysis was conducted in order to determine the frequency with which the peer comments regarding ELLs fall into the seven sections of the peer review guidance.

The second and third questions necessitate the use of qualitative data analysis methods. Deductive methods were used, with the seven sections of the peer review serving as themes for coding purposes. In addition, inductive methods were used to determine additional themes that do not easily fall into one of the seven already established themes from the peer review guidance.

Quantitative Data Analysis

Quantitative data analysis is the application of a numerical representation to a situation. Gall, Gall, and Borg (2003) suggested that quantitative analysis is an attempt to reduce a situation to its most basic constituent parts. In the current study, the amount of data is substantial. For each state, the review comments consist of approximately ten to thirty pages of comments per state included as margin notes on the peer review forms. Quantitative analysis was useful to hone in on the important issues to ELLs in this large data set. In addition to the review comments themselves being substantive, there are also fifty states' worth of data included in this set. Quantitative analysis is important in order to determine which issues in the data are the most relevant, and which issues are unique to a single state.

In the case of this study, a matrix was created to determine the frequency with which issues for ELLs are raised within each of the sections of peer review. Data was analyzed for each state, so that states could be ranked according to those with the most mentioned issues as well as those with the fewest mentioned issues. In addition, Critical Element sections were analyzed and rank ordered, according to those sections with the most comments relevant to ELLs and those sections with the fewest relevant comments. Data

was examined holistically without noting the names of states in the rank ordering; data was also considered with the state labels attached.

The quantitative analysis allowed for a big picture view of the issues relevant to ELLs that were included in the peer review comments. The quantitative analysis allowed for an understanding of which Critical Elements are most salient to assessing ELLs, and which Critical Elements are of lesser importance. The qualitative analysis allowed for more detailed insights into the specific issues that can be addressed in order to improve assessments for ELLs. In addition, the qualitative analysis uncovered the recommendations for best practices mentioned by the peer reviewers.

Qualitative Data Analysis

Qualitative data analysis involves the exploration of data for relevant themes following a variety of possible procedures. The determination of the most appropriate analytic procedure can be facilitated by considering two dimensions: 1) whether the themes were created a priori or emerged during analysis, and 2) the level of complexity that is needed for the analysis (Tashakkori & Teddlie, 1998).

Descriptive content analysis is an appropriate approach for this study because this method “takes a volume of qualitative material and attempts to identify core consistencies and meanings” (Patton, 2002, p. 453) by

identifying, coding, categorizing, classifying, and labeling the patterns of meaning in the data.

In the case of this study, qualitative analysis was useful in order to determine issues that were important across the seven sections of the peer review. Some issues may have been identified as key issues, but were not considered on the same level as the seven sections. Furthermore, within one section, there may be several key issues, and qualitative analysis can be useful in uncovering these. For example, accommodations is an issue that is of importance to the assessment of ELLs. However, accommodations is a topic mentioned directly in the critical element of technical quality. Thus, the thematic analysis did not attempt to connect the themes to a Critical Element, but instead, looked for themes holistically throughout the comments.

Qualitative analysis was also used to uncover other themes that may have arisen during the peer review process that are not directly mentioned in the guidance document itself. In addition, qualitative analysis was useful to find patterns in the reviewers' recommendations, which are separate from the peer review sections. These best practices may not be requirements of the law, but instead, issues that the reviewers thought were important to raise. Thus, the qualitative data analysis is an expansion of the quantitative analysis that brings additional meaning to the research results.

Sources of Data

Data for this research project comes from two sources: comments from the peer reviewers are the primary source of data, and state assessment decision letters are a secondary data source. Each of these types of data will be described in more detail in the following paragraphs.

Peer Review Comments.

All states have submitted evidence to demonstrate compliance with the standards and assessments components of NCLB. These materials have been reviewed in order to determine whether or not a state's standards and assessment system meets the meets the requirements of NCLB. The peers are an anonymous group of state agency personnel and specialists in the field assessment, including those familiar with second language acquisition. During the review process, peer comments are recorded anonymously, both in the margins next to relevant sections, and in summary form at the end of each section. These comments are then shared with the state.

All peer review comments as of December 2006 were reviewed. By this point, a few states had received approval for their systems and had one set of comments. Other states were in the process of submitting additional evidence for review. Still others had more than one set of comments because they had been through more than one round of reviews. However, the peer review

process had become less formal, and some information was not documented in writing. However, key issues for ELLs were identified for all states by December 2006. All rounds of peer review comments available by December 2006 were included as a basis for analysis to aid in identifying themes and examples of both acceptable and insufficient evidence of those themes. The peer review comments were released to the researcher by the U.S. Department of Education in November 2006, with additional comments released in January 2007. The comments are used here with the permission of the U.S. Department of Education.

State Decision Letters.

After the peers reviewed the state's evidence, the U.S. Department of Education made an official determination about the status of the state's standards and assessment system. These determinations are found in the official letters from the U.S. Department of Education to the states. States may receive full approval or approval pending if there are only minor issues to address. If a state has outstanding issues such that the Department of Education cannot grant approval pending, a state may receive a determination called conditional or mandatory oversight. In these cases, states may have some or all of their Title I funds withheld. The letters are publicly available on the ED.gov website.

In addition to letting states know the status of their review, the letters also inform states that have not been approved of the key issues they must address in order to have their system fully approved. This information is useful when compared against the peer comments because the letters help the states understand the relative weight of the individual comments. Thus, these letters served as a secondary data source in that they aided in the interpretation of the peer comments.

The outcomes of the states' peer reviews were compiled in a table according to the name of state and the status of the peer review as approved or not approved. States that were approved after the December 2006 date are noted. Figure 3.1 below illustrates the number of states that were approved by December 2006 (n= 18), the number of states approved after December 2006 (n= 22), and the number not yet approved (n = 9). The approval status for one state, Texas, is unclear according to the publicly available decision letters available on the ed.gov website, and in a follow up conversation with a colleague at the US Department of Education, the researcher learned that the Department of Education is behind schedule in posting letters to the public website. See Appendix B for a complete listing of the approval status of states as of December 1, 2009, according to publicly available information.

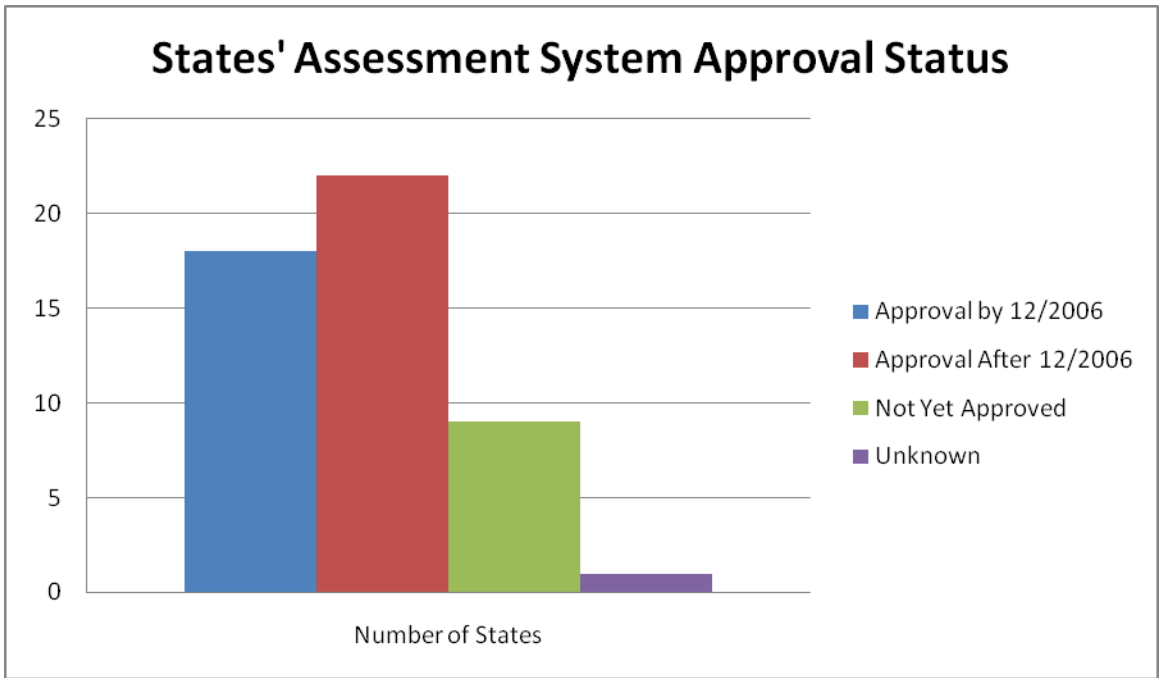

Figure 3.1 States' Assessment System Approval Status as of December 1, 2009

Analysis Procedures

The first phase of the data analysis involved the development of the matrix of comments according to both states and sections. This analysis was used to determine the frequency with which peer comments were made according to each section and according to states. States were rank ordered according to the states with the most comments relevant to ELLs and those with the fewest comments relevant to ELLs. In addition, comments were ranked according to the seven sections of the peer review in order to determine which sections are ranked as having the most comments relevant to ELLs and which are ranked as having the fewest relevant comments. In the second phase of data analysis, an analysis for themes that were established a priori from the peer review was conducted, with the seven categories of critical elements serving as the initial themes for analysis.

The next phase of data analysis served to determine themes that emerged during the analysis process. In order to establish reliability of the qualitative analysis, a protocol was established (Creswell & Plano Clark, 2007). This protocol involved copying each electronic copy of the peer review reports and placing them in a file marked with the state name. After the data set was compiled, each peer review file was handled according to the following process:

1. All peer review comments for a state were read thoroughly.

2. All comments relevant to ELLs were highlighted directly in the report.

3. Margin notes were used to establish topic markers for the themes

4. Highlighting flags were used to note where all comments were located.

5. Notes, including researcher reflections, were jotted down on the outside of the file folder for each state.

6. The topic markers for each state were noted on the outside of the file folder.

7. A table was created in a word processing program, with each state's name and one column for the themes for each state.

8. After all states had been read, and topic markers were established, information was entered into the table. As information was entered, the files were re-read for accuracy.

9. Five states' data were double checked by a colleague. The colleague read copies of the state profiles and was asked to identify themes. Inter-rater reliability was high, with the researcher and her colleague agreeing on all but two coding examples. These two items were discussed, and agreement was established.

9. A word processing program was used to assist with grouping of

themes and establishing the frequency counts for themes.

This protocol was followed to ensure the fair and equal treatment of the qualitative data.

As explained by the protocol, issues were determined to be themes if they appeared in the comments of five states, or ten percent of all states.

Issues relevant to English language learners were flagged and coded. Initially, all statements relative to ELLs were identified. Topical markers for these issues were determined inductively, and a coding book was established. Codes were grouped according to themes. Thus “monitoring accommodations” and “providing a list of sufficient accommodations” were grouped together under the theme “accommodations.” The themes were determined to be necessary and sufficient because they encompassed all of the identified comments in the data set.

Issues were determined to be relevant to ELLs if ELLs were mentioned directly in the comments. Additional comments were also seen to be relevant to ELLs even when the comments were more indirect. Phrases including “special populations,” “subpopulations,” and “diverse stakeholders” were determined to include ELLs. Comments about special education or students with disabilities were not included unless ELL/LEP students were also mentioned directly. Comments relating to parents in general were considered

to be relevant. Assessment practices, such as bias review, DIF analysis, and studies of consequential validity, that address the needs of marginalized groups of students were included in the analysis. However, assessment practices in general, such as standard setting and alignment studies, were only included in the analysis when ELLs were mentioned directly in connection to these practices.

Best practices mentioned by the reviewers were identified. These practices are those additional activities suggested by the reviewers that states can do to improve their assessment systems. That is, best practices are those practices that are above and beyond the requirements of the peer review. For example, a peer reviewer may have suggested a state develop a native language assessment when this type of assessment is not required by federal law. Or, a reviewer may have suggested that a state conduct a research study on the effectiveness of accommodations. Such a study is not a requirement of NCLB, but is considered to be a best practice. In addition, best practices differ from themes in that they did not need to be mentioned in multiple states' peer comments in order to be included in the analysis.

Validity Check

Valid interpretation of research results involves triangulation of data, clarification of researcher bias, and the reporting of all research results, no

matter how well the results fit with the research question or hypothesis (Creswell and Plano Clark, 2007).

The initial entry of the data reporting the distribution of comments across the Critical Elements was completed by the researcher. To check the data, a colleague familiar with the peer review process examined the peer review comments of ten percent of the data (n= 5 states) and completed an independent review of the comments. There was full inter-rater agreement between the researcher and her colleague.

The preliminary results of this analysis were shared with the U.S. Department of Education assessment team as a check on the themes that were derived and as a face validity check on the best practices identified in the comments. This validity check included a phone meeting with a key member of the assessment team that supported the peer review process by convening peer reviewers, and communicating information from reviewers to the states. Through the phone meeting, the findings of this study were discussed and confirmed. One theme was called into question, and upon further discussion, it was decided that this theme, alignment, should be dropped from the analysis.

Summary

Mixed methods were chosen as the research approach to this study

because it is the most appropriate given the research questions. Quantitative analysis will be used to determine the frequency of peer reviewer comments in relation to both states and peer review sections. Qualitative analysis will be used to uncover themes, both deductively as they relate to the seven sections of peer review guidance and inductively as they emerge during the analysis. The next chapter will discuss the results of this data analysis.

CHAPTER IV: RESULTS

However beautiful the strategy, you should occasionally look at the results.
--Winston Churchill

This chapter reports the research findings from the study. The data collected for the study included the frequency and distribution of comments about issues important to assessing ELLs in each of the Critical Elements of the Standards and Assessment Peer Review. In addition, the peer comments were analyzed thematically for the key issues relevant to ELLs that states need to address for the peer review as well as best practices identified by the peer reviewers.

The research questions for this study include the following:

1. What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be in compliance with the standards and assessments accountability requirements of No Child Left Behind?

2. What are the commonalities among states in meeting the federal requirements for assessing English language learners?

3. What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest?

Questions 1 and 2 are answered by both the quantitative data and the qualitative data. First, the frequency and distribution of peer comments are reported. Then, the themes that emerged from the qualitative analysis are considered. Question 3 is answered by the qualitative analysis.

Findings

As of the end of December 2006, there were a total of 18 states that had their assessment systems approved through the peer review process. The remaining 32 states were sent letters detailing elements that remained outstanding. Of these 32 states, all but one state's letters included issues relevant to ELLs that must be addressed in order for the state's assessment system to be fully approved (see Table 4.1). This is according to information that is publicly available on the ed.gov website. However, the Department of Education has acknowledged that it is behind in sending letters and posting them to the website, so the actual number of approved states may be much higher than what is available on the website.

Number of States with Approved Assessment Systems Through December 2006	18
Of States with Unapproved Systems, Number **with** Comments Relevant to ELLs in Assessment Letters	31
Of States with Unapproved Systems, Number **without** Comments Relevant to ELLs in Assessment Letters	1

Table 4.1 Comments Relevant to English Language Learners in Decision Letters

As mentioned in Chapter 3, the decision letters provide an explanation of the issues that states must address as an absolute priority in order to have their assessment system approved. In contrast, the peer review comments include these absolute issues as well as suggestions for best practices. Thus, a state may have received full approval and could still have comments relevant to improving assessments for ELLs. A total of 50 states had comments relevant to ELLs in their peer review comments. That is, all states were given identified suggestions for making their assessment systems better for ELLs. The remainder of this chapter presents findings from the peer comments.

Quantitative Results

Question 1: What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be in compliance with the standards

and assessments accountability requirements of No Child Left Behind?

To answer this question, a matrix showing the frequency and distribution of peer comments across states was created. States were considered to have issues relevant to ELLs in a Critical Element when ELLs were mentioned in the peer notes for that Critical Element, and was marked as "yes." If comments pertaining to ELLs were not mentioned for a Critical Element, the state was determined to not have a relevant issue in that Critical Element. In this case, the state was marked as "no." The complete matrix showing all states and all Critical Elements is in Appendix C.

In general, there were more cells in the matrix marked as relevant to ELLs (n= 212) than there were cells marked not relevant to ELLs (n=138). In addition, there were comments relating to the assessment of ELLs in each of the seven Critical Elements of the peer review guidance. The distribution of comments across the Critical Elements is shown in Figure 4.1. In this figure, the y-axis indicates the number of states that had peer comments in each Critical Element.

As illustrated by Figure 4.1, the Critical Element with the most comments relevant to ELLs was Critical Element 4: Technical Quality, with comments in 47 states' peer reviews. Critical Element 6: Inclusion was the

second highest, with comments in 42 states' peer reviews. Critical Element 7: Reporting followed closely with comments in 39 state reviews. Critical Element 1: Content Standards and Critical Element 2: Achievement Standards were tied with comments in 33 states' peer reviews. Critical Element 3: Annual Assessment followed with comments in 15 states peer reviews. Finally, Critical Element 5: Alignment had the fewest number of comments in state peer reviews, with 3.

In doing policy work across states, issues may be seen to be relevant if they are mentioned in the context of ten percent of the states, or 5 states (see Christensen, et al., 2008). It is the case that the issues of Technical Quality, Inclusion, and Reporting are relevant for the vast majority of states. Similarly, issues with regard to Content Standards and Achievement Standards are important for many states. However, although Annual Assessments has many fewer comments, it is nonetheless relevant for at least ten percent of states.

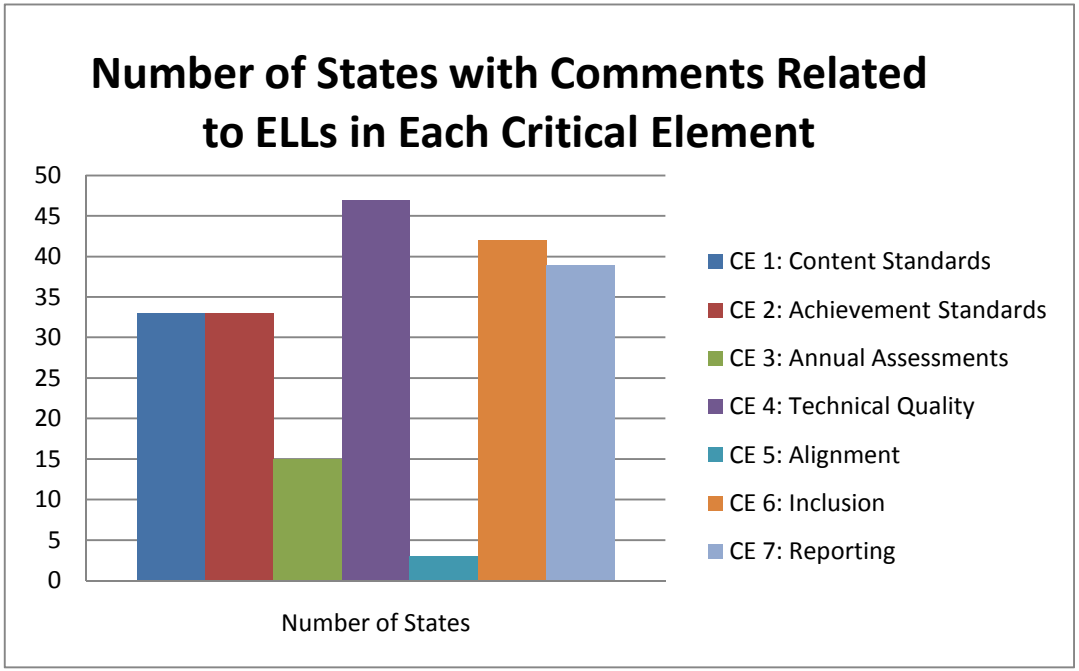

Figure 4.1 Number of states with comments related to ELLs in each critical element

As mentioned previously, all states had at least one comment relevant to ELLs in their peer review comments. The frequency of comments was determined by the distribution across Critical Elements. To account for individual differences across review panels, the actual number of comments were not tracked, but instead, the number of critical elements receiving comments were recorded. Therefore, the highest number of comments a state could have is seven, and the lowest number of comments a state could receive is zero. No state received no comments on ELLs in their peer review. Only four states (Alaska, Maine, Mississippi, and Oregon) had comments in two or fewer Critical Elements. Most states had at least three or more comments relevant to ELLs in their peer comments. See Figure 4.2 for more information on the frequency of comments on ELL issues by state. Almost half of the states had as many as many as five (n= 13 states) or six (n= 11 states) Critical Elements with comments. No state received comments in all seven Critical Elements.

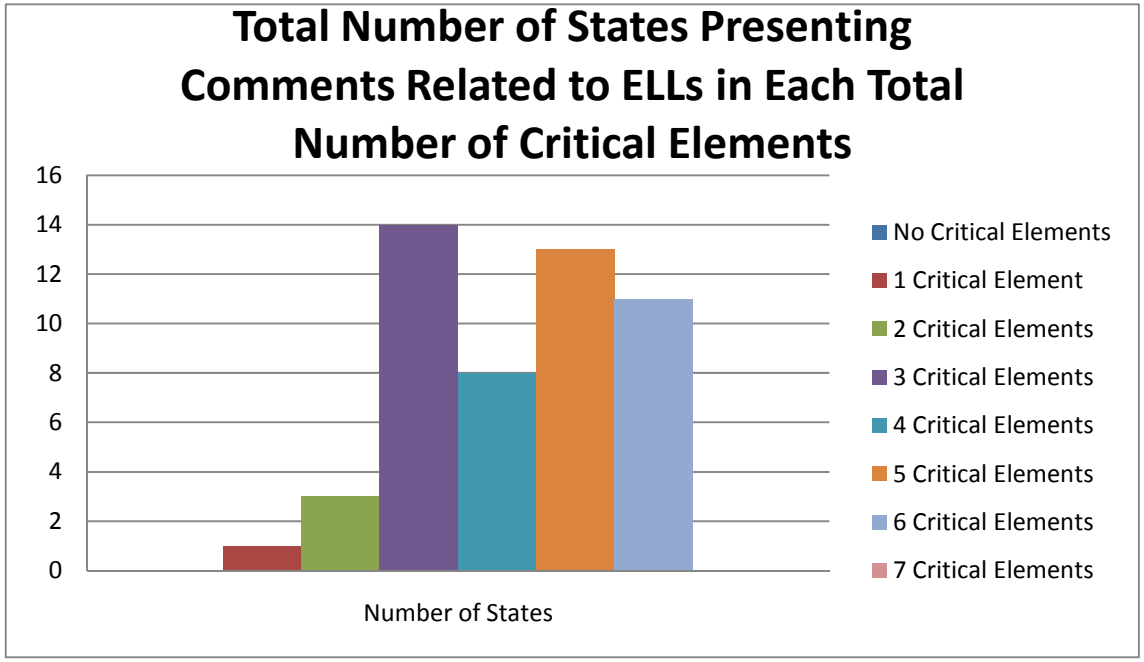

Figure 4.2 Total Number of States Presenting Comments Related to ELLs in Each Total Number of Critical Elements

The states that received the most comments received comments in six of the Critical Elements. These states (n=11) were the following: Arkansas, Iowa, Kansas, Michigan, Minnesota, Nebraska, New York, Texas, Virginia, and Wisconsin. Of these states, three (Arkansas, Iowa, and Michigan) received full approval of their assessment systems by December 2006.

Returning to research question 1, What are the key issues for the assessment of English language learners that are raised from the peer review process in order for states to be in compliance with the standards and assessments accountability requirements of No Child Left Behind? According to the frequency and distribution of peer comments related to ELLs, the primary issues that states must address are those of technical quality, inclusion, and reporting. In addition, the other major areas of the peer review, including establishing content standards and achievement standards, maintaining annual assessments, and ensuring that assessments are aligned to standards are also relevant to ELLs, as established by the frequency of comments about ELLs by the peers in these areas. The analysis of frequency and distribution of peer comments says something about major areas of concern for including ELLs in assessment systems. However, to understand better what the specific issues are, it is necessary to turn to the thematic analysis of comments.

Qualitative Results

In addition to determining the frequency and distribution of peer comments across states, the peer comments were also analyzed thematically. A total of twelve themes were uncovered in the data. They include the following: Accommodations (n = 42 states), Diverse Stakeholders (n= 41 states), Reporting Accuracy (n = 28 states), Inclusion (n = 25 states), Reporting for Parents (n = 20 states), Migrant Students (n = 19 states), Comparability of Test Scores for ELLs (n = 15 states), Participation of ELLs (n = 12 states), Bias Review (n = 10 states), DIF Analysis (n = 10 states), and Standards (n = 9 states). As mentioned in Chapter 3, themes were identified as relevant to the analysis if they were mentioned in the policies of at least five states, or ten percent of all states. Each of these will be discussed in more depth in the following pages in order to answer research question 2.

2. What are the commonalities among states in meeting the federal requirements for assessing English language learners?

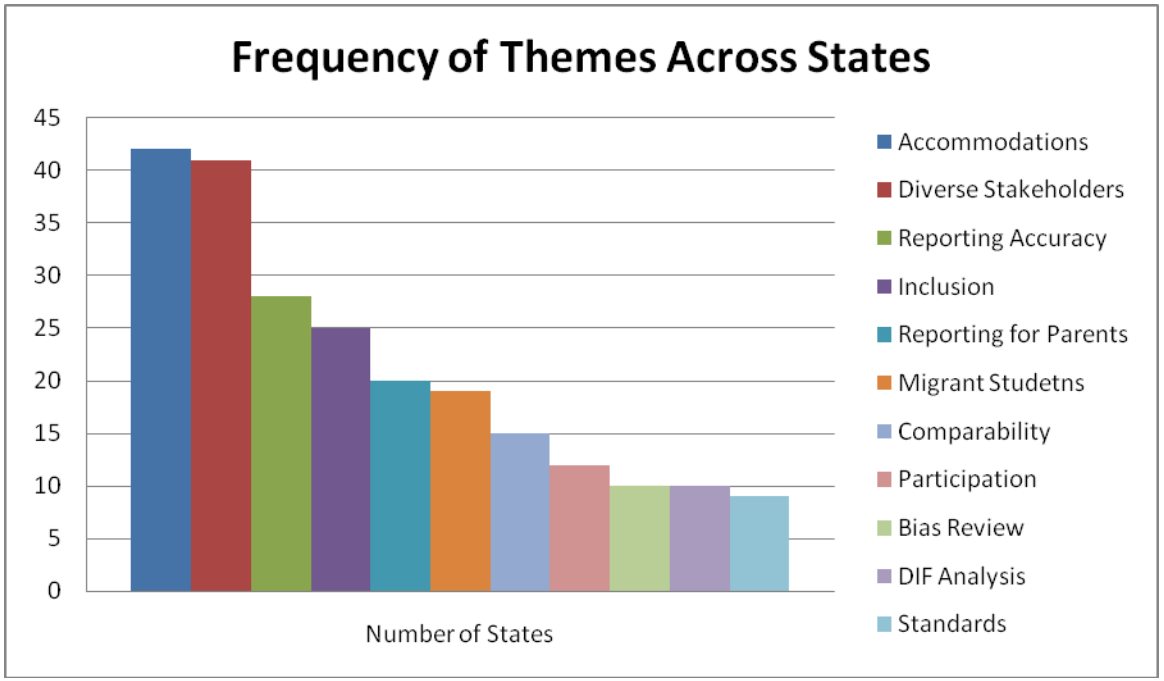

Figure 4.3 Frequency of Themes Across States

Accommodations

Accommodations are changes to instruction or testing procedures that allow students with disabilities and ELLs to access the general curriculum and to show what they know and can do on assessments. In the peer review comments, accommodations was the most frequently mentioned theme, discussed in the comments of 42 states.

There were several key concerns mentioned by the peer reviewers about accommodations. One concern was that ELL students have fewer accommodations from which to choose than do students with disabilities. For example, in Rhode Island, the peer reviewers also observed a limited number of accommodations for ELLs:

> The states need to review the needs of the LEP population and make available a broader range of linguistic accommodations. The states should examine the research in this area. The states need to have policy statements as to why linguistic accommodations are or are not provided.

In addition, in Massachusetts, a recent law eliminated the use of translated assessments, and the peer reviewers noted that translated assessments had been a primary accommodation for ELLs in that state.

Another concern about accommodations was ensuring that those

accommodations used by an ELL were also available to the student for instruction. In addition to being a best practice, NCLB requires that states ensure that assessment accommodations are also used by the student for instruction. In West Virginia, for example, the peer reviewers stated, “The state should be asked to revise its procedures to specify that LEP and other accommodations used for assessment must be restricted to those used for instruction.”

Monitoring accommodations is another issue that was raised by the peer reviewers. Monitoring accommodations may involve coding information on test forms, in-person observations during test day, and record reviews. Goals of monitoring include ensuring that students get the accommodations they are supposed to use, and that these accommodations are ones that have been used by the student in instruction. Louisiana submitted some information to show that they monitor accommodations. However, the peer reviewers had additional comments and suggestions for this state:

> The primary means of monitoring accommodations provided to students (SWDs and LEP students) appears to be in the coding of answer documents. The extent to which this information is summarized and used is not clear/explicit in the documentation provided. The Peers recommend that Louisiana expand its monitoring checklist to include

> more specificity with respect to the types of accommodations provided students.

Thus, Louisiana was asked to provide additional information on how they use the data they do collect to monitor accommodations, and at the same time, the reviewers suggested additional best practices to include expanding the state's monitoring checklist.

An additional area of concern within the topic of accommodations was regarding the validity of accommodations used for ELLs. Several states were asked to provide additional information to demonstrate that accommodations used for ELLs result in valid and meaningful scores. In North Dakota, the reviewers recommended using an expert panel to review accommodations:

> The State should consider using expert panels to review the use of accommodations on the NDSA and to review available research in order to confirm that the use of accommodations yields meaningful scores. If the State has already done so, it should provide appropriate documentation.

Expert judgment informed by research is a common means of validating the use of accommodations given that there currently is little consensus on the use of specific accommodations. In other cases, the validity of a specific accommodation was called into question. For example, Missouri allows the use

of the read aloud accommodation, which means that ELLs can have items from the test read aloud. The peer reviewers wondered if this was a valid accommodation for the test, and they requested that the state provide evidence on the validity and reliability of this accommodation for the assessment.

Diverse Stakeholders

Diverse stakeholders refers to the importance of having people who understand the issues of language acquisition and cultural competency involved in the development of standards and assessments. The theme of diverse stakeholders was mentioned in the peer review comments of 41 states.

Concerns about the representation of ELLs in the stakeholder groups of both the content standards and achievement standards development groups were mentioned. For example, although California was commended for having a broad representation of stakeholders in developing its content standards, the peer reviewers noted the following with regard to the development of achievement standards for the alternate assessment:

> the educators and community members used were predominantly white females. The diversity of the groups was quite limited, especially for a state as diverse as [California].

In other states, diverse stakeholders were lacking in the development of both content standards and achievement standards. For example, with regard to the content standards, peer reviewers for Colorado noted, “There is no evidence of the involvement of educators with expertise in the education of students with disabilities or English language learners, parents, citizens, and business leaders.” For the development of achievement standards, the peer reviewers also observed that “the available evidence indicates a lack of wide involvement of stakeholders in the creation, review, and validation process.”

Reporting Accuracy

States are required to report the number of students who took an assessment. Reporting accuracy was an issue mentioned in the peer comments of 28 states. For some states, the issue of accurately reporting the number of ELLs who were tested and not tested was an issue of concern. For example, in Arizona, the peer reviewers commented the following:

> State policy requires participation of all students. However, evidence of the inclusion of all students is not provided. Statistical reports seem to report that more than 100% of the eligible students are tested. Policy calls for the inclusion of all students but test information calls this into question.

Other states had similar questions, such as Maryland, where peer reviewers

noted that there was “No evidence about participation rates and assessment results for subgroups.” In Mississippi, the state was asked to provide additional information on “enrollment numbers of students by grade level and content area ***or*** the number of students tested/not tested by grade level and by content area.” In some cases, these requests for additional information were general, applying to all students. In other cases, there were notations that not all subgroups, including ELLs, were reported. However, it should be noted that in most cases, the omission of this information was related more to lack of good record-keeping or failure to submit information rather than an attempt to be misleading about the number of ELLs tested.

One example of this ambiguous information comes from Pennsylvania. In this state, the peer reviewers could not clearly determine the accuracy of reports from the information given, as stated in the following:

> It is not clear from the State Report Card 2004 – 2005 which students are included and not included. No reporting is included below the State level that shows the actual students enrolled and the proportion represented in the reports.
>
> What has been done to ensure that these reports are accurate and represent all of the students who should be assessed in each school and district?

It is likely that in future rounds of the review, Pennsylvania was able to submit additional information to satisfy these questions on the part of the peer reviewers.

Another reporting issue that came up for some states was that of the minimum n-size. Some reviewers suggested that a state's minimum n-size was too big, and other states were told their minimum n-size for reporting a subgroup was too small. For example, Minnesota reporting subgroups with as few as 2 students. The peers responded that

> A minimum reporting size of 2 seems too small to maintain confidentiality. Evidence on how minimum reporting size is actually set for various reports would be helpful in understanding how MN ensures that no personally identifiable information is reported**.**

Subsequent to the peer reviews completed in December 2006, the federal government has released guidance that requires states to maintain the same minimum n-size for all subgroups (US Department of Education, April 2007).

Inclusion

Inclusion refers to the requirement that all students have access to grade level curriculum and are included in the assessment and accountability system. States are required to have policies stating that all students will participate in the assessment and accountability system, and states are also

required to submit evidence that the policies are implemented. Inclusion was an issue mentioned in half of the states' peer reviews (n=25).

In some cases, the peer reviewers could not ascertain whether or not the state had a policy on inclusion. For example, in Minnesota, the peer reviewers asked, "Does MN have a formal policy on the inclusion of all students in state assessments? What evidence is there that a formal policy that has been approved by the state?"

In other states, there was concern by the peer reviewers that the state's policies on inclusion were being implemented. In one state, Wyoming, the peers were very direct in their observation: "WY has a policy regarding inclusion but provides no evidence that it is being implemented." However, in other states, the peers were more nuanced, observing several aspects of inclusion that needed to be addressed by the state. For example, in Colorado, the peer reviewers wrote the following:

> Inclusion of all students: Need data on number of students enrolled vs. number of students assessed, by grade level, by content area, by subpopulation group. Need Test Administration Manuals to document the implementation of policies and procedures regarding the assessment of students with disabilities, ELL students, and mobile students. Need evidence that Colorado monitors the degree to which

its stated polices and procedures are being implemented.

This particular example illustrates the close relationship among inclusion, participation, and reporting accuracy.

Finally, there were some states that did not demonstrate through their reports that ELLs were included appropriately in the assessment system. For example, Connecticut received the following comments from the reviewers:

> The reports provided do not show that the ELLs are included in assessments to the degree that they should be; and
>
> For 10th grade, it is not clear that ELLs, Black Students, Hispanic Students, and Economically Disadvantaged students are included in the assessments to a sufficient degree.

Thus, it is not enough to have a policy on inclusion, but it is also necessary to demonstrate through assessment reports, that all students are included in the assessment system.

Reporting for Parents

Reporting information on assessment participation and performance for all students in an accessible format is a cornerstone of the accountability requirements of NCLB. States are expected to provide assessment reports in a format accessible to parents. In this particular theme, reporting for parents of ELLs was rarely mentioned directly in the peer reviewers' comments.

However, states were counted as having relevant comments when parents in general were mentioned, which was in the comments of 20 states.

In some cases, information on reports for parents was assumed to be missing. In Oregon, the reviewers noted that a report for parents should be "developed (or forwarded to the US Department of Education)." Another state, Pennsylvania, also did not provide clear information about when and how reports are sent to parents. In addition, in this case, the peer reviewers also noted missing information on how many parents make use of online reports:

> [Pennsylvania] provides materials and reports by way of the WWW for teachers to help interpret results to parents and for parents to get information on how to help their children in areas of low performance. No evidence is provided on how many PA parents or teachers have access to and/or make use of the materials.

The peers noted that materials must be available to parents in accessible formats, such as the internet, but it is important to keep track of how many parents are accessing the materials in this format.

One other issue relating to reporting for parents is the clarity of the reports themselves. The peer reviewers noted that assessment language is often full of jargon that must be explained for a layperson such as a parent. In addition, reports must not be too general. For example, in Texas, the peer

reviewers observed that “the objectives and the descriptors in the Parent Guides for the [Texas general assessment] do not always vary by grade level. This is potentially confusing for parents who have children in different grades.” Thus, it is not enough to supply reports to parents; these reports must be in accessible language and appropriately targeted to each child’s grade level.

Migrant Students

Migrant students are students who work in seasonal agriculture or have an immediate family member who works in seasonal agriculture. Migrant students may or may not be ELLs; however, many migrant students are ELLs, and thus this theme is relevant to the study at hand. Issues concerning migrant students were mentioned in the peer review comments of 19 states.

The primary issue regarding migrant students is ensuring that migrant students are included in the accountability system, and providing evidence to support their inclusion. For example, in Florida, evidence of including migrant students in assessments was not provided to the reviewers. Similarly, in Illinois, reviewers noted that “Evidence is needed of policies and practices of identification and inclusion of migrant and other mobile students.” States may have policies and practices to include migrant students in the assessment system, but they did not submit evidence to support what they do.

Comparability

One common theme in the data was that of comparability of alternative forms of the general assessment. Not all states have alternative forms of the general assessment, however. Figure 4.4 illustrates the number of states with no native language assessments (n= 29), native language assessments (n=17), and simplified English assessments (n= 5). In some states, native language assessments are offered. In other states, they may not be offered. Some states offer on the spot translation of the test into the student's native language. One state refers to the version of the general assessment as an

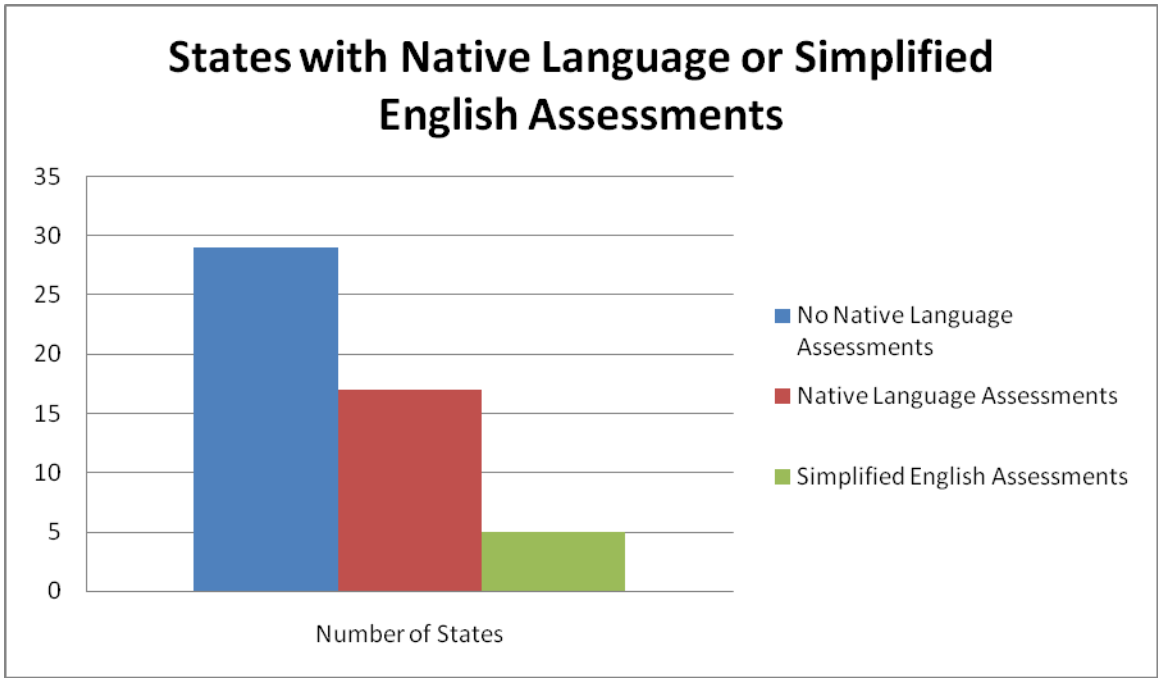

Figure 4.4 States with Native Language or Simplified English Assessments

alternate assessment for LEP students. Thus, the reviewers wanted to ensure that the test is comparable to the general assessment and not the alternate assessment for students with the most significant cognitive disabilities. In all, fifteen states had issues relating to comparability of test formats for ELLs in their peer review comments.

Some states had issues with comparability of translated assessments. In Delaware, for example, the peer reviewers asked the state to provide additional information about the translation process: "There is a process in place for translating the tests. However, there is no evidence that the results obtained on the translated tests are in fact comparable. Backward translations may not be sufficient for showing comparability." In Maine, local translations are used. This means that on the day of the test, an interpreter is provided to translate the test directly for the student. The peer reviewers raised concerns about this process as well. They stated:

> Local translation of tests and/or responses would typically invalidate the test results. The burden of proof resides with Maine. Maine should validate the translation process or discontinue it because although it may decrease systematic error, it will likely increase random error.

Although translated assessments pose challenges for states in ensuring that scores from these alternative assessments are comparable to scores from the

general assessment, states that did not have a translated assessment often received peer comments suggesting that they develop a translated assessment. This was the case for Arizona, for example, which is an “English Only” state and consequently does not offer translated assessments.

States with simplified English assessments also raised concerns for the peer reviewers. In Illinois, the reviewers were uncertain that the simplified English assessments were measuring grade level content, and they asked the state to show evidence that these assessments were comparable to the general assessment. Such evidence could include any of the following, according to the reviewers: “alignment data, content expert judgments of comparability of content, expert judgment of and/or statistical and empirical data showing the comparability of scores [especially at the proficient/not proficient cut].”

Participation

Participation is a critical component to the accountability system. When all students participate in assessments, stakeholders, including parents, teachers, and administrators, are able to know what students know and are able to do. When all students do not participate in assessments, stakeholders are unable to get an accurate picture of how a school is performing. One aspect of participation is ensuring that all students participate in the

assessment. Another aspect is ensuring that students participate appropriately. That is, students who need accommodations should be able to use them, and students with significant cognitive disabilities should be able to take an alternate assessment. A total of 12 states had comments related to participation of ELLs in assessments.

In one state, Utah, the reviewers observed that information on participation was missing for some subgroups of students, including ELLs. The reviewers noted that they, "were unable to determine if participation data are available for migrant and homeless students, students with limited English proficiency, or the number of excluded students." Similarly, Florida received similar comments, although in this case, the reviewers also added that information was not available to demonstrate that only students with significant cognitive disabilities were given an alternate assessment:

> The state needs to provide evidence to support its assertions that all students participate in its assessments, that it has guidelines for assessing the students with disabilities with only the most severe disabilities in its alternate assessment, that English language learners are assessed in a manner to best show their achievement (especially for students such as Hispanic students) and that migrant students also participate in its assessments.

States need to have solid policies and practices in place to ensure the participation of all students, including ELLs, in the assessment system.

Bias Review and DIF Analysis

Bias review and DIF analysis are both processes used to ensure that assessments are fair to everyone. With a bias review, a committee comprised of diverse stakeholders meets to review items for bias. In a bias review, the committee is considering more than linguistic bias, but cultural bias as well. DIF analysis is a psychometric process used to determine whether or not some demographic groups tend to perform better than others on certain items. If that is the case, items that favor one demographic group over another will be thrown out. With regard to both bias review and DIF Analysis, these processes were mentioned in connection to ELLs in 10 states' peer review comments, respectively.

When states received comments from the peers about bias reviews, it was because the state did not provide information on the composition and training of the bias review panel. As a result, the peers were not clear about the diversity of the bias review panel, as was the case, for example, for two states, Maine and New Hampshire.

When the peer reviewers made comments about DIF analysis, it was usually because the state did not provide evidence that DIF analyses had been

conducted. This was the case for Washington and Oregon, for example. In other cases, the state provided information on DIF analysis, but did not present information that ELLs were included as a demographic group in the analysis. In North Dakota, for example, the reviewers noted that the state did not provide evidence that ELLs were included in DIF analyses, although the reviewers speculated that the population of ELLs in North Dakota may be too small to include.

Standards.

Standards is a key cornerstone of NCLB. All states are required to develop standards in English language arts, math, and now science, and to assess based on the standards. Standards was mentioned relative to ELLs in 9 states' peer review comments. A key issue for ELLs is that in some states, ELLs are given an alternative assessment for English language arts. In this case, the peer reviewers wanted to ensure that students were tested on grade level standards. This is in contrast to alternate assessments for students with significant cognitive disabilities, who are tested on alternate achievement standards.

This issue was of particular relevance, at the time of the peer review in 2006 because there was some confusion about whether or not ELLs could substitute the English language arts assessment for an English language

proficiency assessment. In a letter addressed to states on February 20, 2004, Secretary of Education Rod Paige explained the testing requirements for LEP students:

> For purposes of participation in a State's assessment system, LEP students, during their first year of enrollment in U.S. schools, must take an English proficiency assessment [§1111(b)(7)], and ***may*** participate in the reading/language arts assessment [§1111(b)(3)].

The letter goes on to state that states are not required to report the English language arts assessment results of first-year LEP students. However, this guidance was somewhat confusing for some states, that thought students could substitute one assessment for another.

As a result, some states suggested that their English language arts assessment was aligned to their English language proficiency assessment. One example is that of Minnesota, which received the following comments from the peers:

> Evidence of English Language Proficiency standards was not submitted as part of the Peer Review documentation. On page 11 of the Peer Review Document English Language Proficiency standards were mentioned, but does MN really have them? If MN has these standards, what is the evidence that the English language proficiency standards

align with the MN Language Arts standards?

In order to provide states with more clarity and guidance around the issue of assessing ELLs in reading and math, the US Department of Education developed the LEP Partnership to provide technical assistance to states on this topic and others relating to ELLs.

Research Question 2 asked, what are the commonalities among states in meeting the federal requirements for assessing English language learners? Through a qualitative analysis of data, the peer review comments show several commonalities among states in trying to meet the federal requirements for including ELLs in the assessment system. These commonalities are accommodations, maintaining diverse stakeholders in the development of content standards and achievement standards, reporting accurate assessment results for ELLs, including ELLs in the assessment system, ensuring that assessment reports for parents are accessible, addressing the inclusion and participation of migrant students, maintaining comparability of test scores from alternative forms of the test, establishing and implementing participation policies that include ELLs, addressing potential item bias and DIF analysis, and ensuring that ELLs have access to the standards.

Research Question 3: What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest?

Research question 3 asked, What recommendations, or best practices, for the assessment of English language learners do the Standards and Assessments peer reviewers suggest? This question was also answered through the use of qualitative data analysis.

The process for the Standards and Assessments peer review is that of a deficit model. That is, the reviewers are looking for information that is missing, and states that pass the review are ones with little to no information missing. The review does not lend itself to noting promising practices on the part of states, nor is it the reviewer's jobs to make recommendations of best practices. Furthermore, the reviewers did not identify best practices in the states. However, within the reviews, there are points in which the reviewers offer states suggestions for how to improve their assessment systems. For the purpose of this study, these suggestions are considered best practices because they are practices that go above and beyond the requirements of NCLB, and were mentioned in the comments of one or more states' peer reviews.

Recommendations

The following are recommendations for including English language learners in the assessment and accountability system, as mentioned by the peer reviewers: (a) attending to the educational needs of linguistic minority students; (b) carrying out studies of consequential validity; (c) providing more alternative forms of assessments, and ensuring that the methods used to develop these alternative forms are valid; (d) researching the validity of accommodations for ELLs; and (e) developing a plan for system improvement.

In at least three states, the peer reviewers recommended that states address the needs of linguistic minorities. Peer reviewers suggested that Alaska develop a research agenda that includes the linguistic minority students as part of their demographics in research. Reviewers suggested to North Dakota that DIF analysis should include Native American populations. Similarly, it was recommended to South Dakota that more be done to include the students of the Hutterite Colony in all aspects of the accountability system.

Studies of consequential validity were another practice recommended by the reviewers. In this case, the reviewers made the suggestion to carry out such studies in the comments of 16 states. Consequential validity addresses both the intended and unintended consequences of the assessment system,

and as a result, is one way for states to ensure that the assessments are having the intended effect of closing the achievement gap for students in the margins, including ELLs.

NCLB requires states to provide a variety of accommodations for ELLs; however, NCLB does not explicitly state which accommodations a state must provide. The peer reviewers made several suggestions to states regarding alternative test formats as an accommodation. They encouraged more states to provide native language tests, and in languages beyond Spanish. They also encouraged the use of video translations to ensure that translations would be fair and equitable to all students using them. Similarly, they suggested that states using local translations should establish policies that explain how to conduct an on-the-spot translation, so that the validity of the test is maintained.

In addition to these alternative test forms as an accommodation, the reviewers recommended that states conduct research on the validity of testing accommodations for ELLs. This research conducted by states can be useful in revising accommodations policies. For example, a research study may help a state determine that a controversial accommodation, such as a bilingual dictionary, may not interfere with the validity of an assessment.

Finally, the reviewers recommended that states develop a plan for

system improvement, and that this plan include ELLs. Oftentimes, when states have developed their assessment systems, they have done so without including marginalized groups, like ELLs and students with disabilities, from the beginning. Unfortunately, states have had to take a more reactive approach, adjusting their policies and practices to become more inclusive. By including ELLs in the plan for system improvement, states will ensure that their plan can be successful from the start.

Research question 3 asked, What recommendations for the assessment of English language learners do the Standards and Assessments peer reviewers suggest? The practices recommended by the peer reviewers include (a) attending to the educational needs of linguistic minority students; (b) carrying out studies of consequential validity; (c) providing more alternative forms of assessments, and ensuring that the methods used to develop these alternative forms are valid; (d) researching the validity of accommodations for ELLs; and (e) developing a plan for system improvement. Again, it is important to note that it was not the role of the reviewers to suggest best practices. These are practices recommended by the reviewers that states might use to improve their assessment systems; these are not practices identified in the materials submitted by states.

Summary

The answers to these three research questions indicate that including ELLs in the assessment and accountability system is very much a part of the standards and assessment peer review. States have genuinely begun to address the inclusion of ELLs, but as the data in this chapter demonstrate, there is much left to do. All states had one or more issues relevant to ELLs in their peer comments. Furthermore, all Critical Elements in the peer review had comments supporting the inclusion of ELLs in the accountability system. The next chapter discusses the findings in further detail. Implications for theory, practice, policy, and research are discussed. Limitations of the current study are described, along with directions for further investigation.

CHAPTER V: DISCUSSION AND CONCLUSION

The important thing is not to stop questioning.
--Albert Einstein

This chapter summarizes the findings from the study and discusses the study's implications for theory, practice, policy, and research. Limitations of the current study are also discussed. This chapter concludes with considerations for further investigation.

Summary of Findings

This thesis began with a discussion of the students who are at the center of this research investigation: the English language learners in U.S. schools who take tests every year to show what they know and can do, both in content areas and in developing English language proficiency. The theoretical underpinning of the accountability system is that through holding states and schools accountable for the education of all children, schools will improve their educational practices, and outcomes, as measured by accountability assessments, will illustrate these improvements. As noted in the introduction to this thesis, English language learners are the fastest growing subgroup in our schools today, and at the same time, as a group, they are scoring below proficient on accountability tests. ELLs must be included in the educational system, and this includes accountability assessments.

This study reviewed the peer comments related to assessing ELLs that were included in the federal Standards and Assessment Peer Review for all fifty states. Comments were analyzed for frequency and distribution across the seven Critical Elements of the peer review. All seven areas had comments in one or more states' peer reviews relative to assessing ELLs. In addition, all fifty states had comments in one or more Critical Elements regarding assessing ELLs. Based on these findings, this study concludes that the topic of including ELLs in the assessment and accountability system is of significance.

Comments were also analyzed thematically, using an inductive analysis. Eleven themes emerged from the analysis, and were verified by a colleague at the U.S. Department of Education. These themes are Accommodations, Diverse Stakeholders, Reporting Accuracy, Inclusion, Reporting for Parents, Migrant Students, Comparability of Test Scores for ELLs, Participation of ELLs, Bias Review, DIF Analysis, and Standards. Themes were established if the topic appeared in more than 10 percent of states, or at least 5 states. These themes appeared in 18-84 percent of states. Thus, these themes can be seen as issues that need to be addressed in order to improve accountability assessments for ELLs.

Ultimately, the lesson learned from analyzing these peer review comments on state standards and assessments is that if ELLs are to be fully

included in the accountability system, there is much work to do. States have begun to make progress toward addressing the inclusion of ELLs, but they must ensure that issues relevant to ELLs are attended to throughout the process of setting standards and developing and implementing assessments based on those standards.

Implications

In Chapter I, the Donovan, Bransford, and Pellegrino (1999) framework illustrating the paths of influence that educational research has on classroom practice was described. This theoretical framework put forth four paths from research to practice: policy, teacher education, educational materials, and the public. The implications from the study for each of these paths is described in more detail below.

Implications for Policy

The findings from this study have implications for educational policy and their underlying theories. As stated in the introduction, accountability theory is based on the premise that school outcomes for students in the margins, including ELLs, should be improved when schools have greater expectations of accountability. ELLs have been previously excluded from assessments, psychometrically—because they are "too diverse" of a population to analyze statistically, and practically—because of potential

concerns about poor performance. When ELLs have been more recently included, their performance has been below proficient.

Using accountability theory, it is the argument of this thesis that ELL performance is low because the tests are not fully accessible to this group of students. Because of this past exclusion, educators and test developers are still working to catch up—developing tests that appropriately include ELLs, with or without accommodations, and providing ELLs with access to grade level curriculum. The issues involved in including ELLs in assessments demonstrate themselves to be persistent in this study—comments relating to ELLs were found in every Critical Element and in all 50 states. As states continue to work to make assessments more inclusive, ELL assessment scores should increase.

Implications for Teacher Education

The research study does not lend itself directly to implications for teacher education. However, a number of the themes that emerged from the analysis are issues that can be addressed through teacher education. These include accommodations, diverse stakeholders, reporting for parents, and standards.

Accommodations is an area in which both pre-service and currently practicing teachers can gain more information. Knowing more about

accommodations available for an assessment, as well as understanding the accommodations decision-making process, can help teachers make sound decisions about accommodations for ELLs. When good accommodations decisions are made, ELLs' test scores are a more accurate representation of their knowledge and skills.

The area of diverse stakeholders was one in which 41 states were noted as not having representatives who were familiar with language acquisition on committees. Pre-service and practicing teachers should receive training on standards and assessments so that they feel comfortable serving on committees that work to develop state standards and assessments.

Reporting for parents is another element that can be addressed through teacher education. When teachers have a strong foundation in assessment and the purposes of assessment, they can be confident in presenting information on assessments and their results to parents. Teachers have more opportunities than many other educators to interface with parents, and communicating positively and confidently about assessments can go a long way in helping a student do his or her best on testing day.

Access to standards and grade level curriculum is another area that can be supported through teacher education. Programs can continue to strive to develop collaborative efforts among teachers, so that ELL specialists and

general educators are able to work together seamlessly to improve outcomes for ELLs. ELLs should be consistently included in the accountability system, through access to grade level, standards based curriculum and participation in assessments, and this mindset can be reinforced through teacher education.

Implications for Educational Materials

In applying the assessment triangle (Pellegrino, et a., 2001) to the development of assessments, state education agency personnel and test developers can take the lessons learned from this thesis in application to both *observation* and *interpretation*. ELLs must be included in the initial design of an assessment, and tasks must be given in such a way that will allow them to show what they know and can do. Accommodations and alternative test forms should be built in from the start whenever possible. Similarly, in making meaning of test results, ELLs should be considered throughout the process. The diversity of this subgroup should be accounted for, so that there is meaning in this subgroup's assessment results. Furthermore, results should be publicly reported and shared with parents in accessible language.

Research on assessment should include ELLs so that research results can continue to inform best practices for including all students in assessments. The themes of accommodations, reporting accuracy, comparability of test forms, participation, bias review and DIF analysis all represent areas of

assessment that could be strengthened by additional research. In particular, the area of accommodations has been researched for students with disabilities, but many fewer studies have considered accommodations for ELLs. Research in this area could inform accommodations decision-making as well as the development of assessments.

Implications for the Public

The public perception of ELLs is oftentimes that they do not want to learn English, although Crawford (2007, 2000) and others have argued against the false nature of this belief. This study does not directly counter this perception; however, it does offer an explanation for lower test performance on the part of ELLs. Although states have made progress in making assessments more accessible to ELLs, there is more that needs to be done before ELLs will be able to show what they know and can do on these assessments. The students, themselves part of the public, may not be surprised by this study's finding that educators can do more to develop more inclusive tests and ensure that students have access to the curriculum. However, this research study also shows that states are improving in their efforts to make tests more inclusive, and ELLs, as members of our society, have the right to education and educational supports. When ELLs participate in assessments, their successes can be documented, and their shortcomings

can be used to continue to improve the educational system.

Limitations of the Current Study

The study undertaken here is not without its limitations. First and foremost, this research study used data from a process that continues to be ongoing. The established cutoff date for data collection was through December 2006; however, the peer review process continues to be underway, and the process has evolved. Additional assessments, including a new alternate assessment based on modified achievement standards, have been added to the review process. States that were once approved may have been changed to an unapproved status due to information that has been introduced since the data was collected. Ultimately, this study represents a snapshot in time.

Another limitation of the study relates to the decision letters, which were part of the process used to validate the information from the peer comments. In a conversation with a colleague at the U.S. Department of Education, it was revealed that the department is experiencing a backlog in posting the decision letters to the web and making them publicly available. Thus, the information I was able to use to verify the relative importance of my themes was only as accurate as the information I was able to gather from the ed.gov website. Some states may have received approval but were not

documented on the department's website.

The peer review process itself is not without limitations. As mentioned earlier, the process operates as a deficit model. Reviewers were encouraged to look for what was lacking in a state's assessment system. They were not encouraged to make note of exemplars. The reviewers themselves brought their own biases and preferences for certain issues to the process. This was particularly evident in one state, in which the review comments mentioned concerns on nearly every page that the state did not offer translated assessments.

In spite of these limitations, the study offers some insights into the review process. State agency personnel can recognize that their state is not the only one grappling with the issues related to including ELLs in assessments. Furthermore, they may benefit by seeing the kinds of issues that were raised in other states, as well as some of the comments made by reviewers. They can be better informed about the process, as it remains ongoing.

Directions for Future Exploration

This study offers a number of directions for future exploration. Given that the peer review is an ongoing process, a follow-up study that looked at data post-December 2006 in comparison to this earlier dataset would be an

important means of documenting both how the process has changed as well as any improvements that have implications for ELLs that have already been made to the assessment system.

In addition, this study represents the perspective of the peer reviewers experience with the process of approving state standards and assessments. Anecdotally, I have heard personnel from several states voice frustration with the process, stating that it is confusing and at times feels unfair. Focusing a study on their experiences with the process through a qualitative interview study could help uncover particularly challenging aspects of the process that could be improved. Furthermore, this is one example of a peer review, and peer review is a standard tool for evaluation used in other contexts. By illuminating the challenges of peer review, other peer review processes could be improved as well.

Conclusion

As mentioned previously, when ELLs have low performance on statewide assessments, there may be a number of mitigating factors, including access to challenging and responsive curriculum and grade level instruction, assessment design, individual student factors, as well as other factors. It is beyond the scope of this thesis to resolve this challenge completely; however, small steps can be made. Through the analysis of peer

review comments from the federal standards and assessments peer review, this study has shown that there are a number of important outstanding issues that must be addressed in order for ELLs to be fully included in assessment systems. For the growing number of ELLs in U.S. schools, having all of these elements—improved access to grade level standards and accommodations, inclusion in the process of developing standards and assessments, tests in alternative formats, assessment reports that are accurate and presented in language that is accessible to parents—may not be enough to ensure the accountability assessments allow ELLs to show what they know and are able to do, but they are a step in the right direction.

.

REFERENCES

Abedi, J. (2007, April). *Current research evidence on the validity of accommodations to help states with the provision of accommodations for ELL students.* Paper presented at the annual meeting for the American Educational Research Association. Chicago, IL.

Abedi, J. (2002). Assessment and accommodations of English language: Issues, concerns, and recommendations. *Journal of School Improvement 3*(1), 83-89.

Abedi, J. (2001). *Assessment and accommodations of English language learners: Issues and recommendations*. (CRESST Policy Brief 4). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.Retrieved February 15, 2009 from http://www.cse.ucla.edu/products/policy/cresst_policy4.pdf

Abedi, J. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Technical Report 478). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved February 15, 2009 http://www.cse.ucla.edu/products/Reports/TECH478.pdf

Abedi, J. (1997). *Dimensionality of NAEP subscale scores in mathematics* (CSE Tech. Rep. No. 428). Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J. & Gándara, P. (2006). Performance of English language learners as A subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 26 (4), 36-46*

Abedi, J.,Herman, J. L., Courtney,M., Leon, S., & Kao, J. (2004). *English language learners and math achievement: A study on classroom-level opportunity to learn*. Los Angeles: University of California: Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Projects Report 1).

Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved March 1, 2009 from http://cehd.umn.edu/nceo/OnlinePubs/LEP1.html

Albus, D., & Thurlow, M. L. (2005). *Beyond subgroup reporting: English language learners with disabilities in 2002–2003 online state assessment reports* (ELLs with Disabilities Report 10). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved February 15, 2009 from http://cehd.umn.edu/nceo/OnlinePubs/ELLsDisRpt10.pdf

Anstrom, K. (1996). Defining the limited-English-proficient student population. *Directions in Language and Education: National Clearinghouse of Bilingual Education,1* (9), 1–9.

Batalova, J., Fix, M., & 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. Retrieved February 15, 2009 from http://www.migrationpolicy.org/pubs/Measures_of_Change.pdf

Bechard, S. (2008). *Hints and tips for addressing alternate assessment issues for peer review.* Paper presented at the Council of Chief State School Officer's Conference on Student Assessment. Orlando, Florida.

Bolt, S., & Roach, A. (2009). *Inclusive assessment and accountability: A guide to accommodations for students with diverse needs*. New York: Guilford.

Castañeda v. Pickard. 1981. 648 F.2d. 989 (5th Cir.).

Christensen, L. L., Lazarus, S. S., Crone, M., & Thurlow, M. L. (2008). *2007 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 69). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 24, 2009 from http://cehd.umn.edu/nceo/OnlinePubs/Synthesis69/Synthesis69.pdf

Christensen, L.L., Lail, K.E., & Thurlow, M. L. (2007). *Hints and tips for addressing accommodations issues for peer review*. Minneapolis, MN:

University of Minnesota, National Center on Educational Outcomes. Retrieved February 24, 2009 from http://cehd.umn.edu/nceo/OnlinePubs/PeerReviewAccomm.pdf

Council of Chief State School Officers. (2009). 2009 National Conference on Student Assessment preliminary conference program. Washington, DC: Author.

Crawford, J. (2007). The decline of bilingual education: How to reverse a troubling trend? *International Multilingual Research Journal, 1*(1). Available online at http://www.elladvocates.org/documents/bilingualed/Crawford_Decline_of_BE.pdf

Crawford, J. (2000). *At war with diversity: U.S. language policy in an age of anxiety*. Buffalo, NY: Multilingual Matters.

Creemers, B., and Kyriakides, L. (2008) *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. New York: Routledge.

Creswell, J. W. & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

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

Cummins, J. (1989). *Empowering language minority students*. Sacramento, CA:California Association for Bilingual Education.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice.* Washington, DC: National Academy Press.

Dunn, W. N. (2008). *Public policy analysis: An introduction*. Upper Saddle River, NJ: Pearson.

Durán, R. P. (2008). Assessing English-language learners' achievement. *Review of Research in Education, 32* (1), 292-327.

Durán, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 573-587). National Council on Measurement in Education and The American Council on Education. New York: Macmillan.

Elmore, R.F. & Rothman, R. (Eds.). (1999). *Testing, teaching, and learning: A guide for states and school districts*. Washington, DC: National Academy Press.

Fry, R. (2008). *The role of schools in the English language learner achievement gap*. Washington, DC: Pew Hispanic Center. Retrieved February 16, 2009 from http://pewhispanic.org/files/reports/89.pdf

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center. Retrieved February 16, 2009 from http://pewhispanic.org/files/reports/76.pdf

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Pearson Education.

Hambleton, R., & Rodgers, J. (1997). *Item bias review*. EricDigest. Retrieved March 1, 2009 from http://www.ericdigests.org/1997-1/bias.html

Kiplinger, V. L., Haug, C. A. & Abedi, J. (2000). *A math assessment should assess math, not reading: one state's approach to the problem.* Paper presented at the 30th National Conference on Large Scale Assessment, Snowbird, Utah, June 25-28.

Kopriva, R. (2008). *Improving testing for English language learners*. New York: Routledge.

La Celle-Peterson, M. W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64*(1), 55-75.

Lau v. Nichols. 1974. 414 U.S. 563.

Lazarín, M. (2006). *Accountability for English language learners in the* No Child Left Behind Act. Washington, DC: National Council of La Raza.

Liu, K. (2009). *Grade-level standards-based science outcomes for English language learners and language minority students: A review of the literature*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Liu, K., & Anderson, M. (2008). Universal design considerations for improving student achievement on English language proficiency tests. *Assessment for Effective Intervention, 33*(3), 167-176.

National Clearinghouse for English Language Acquisition and Language Instruction Education Programs. (2008). *National and regional data and demographics*. Washington, DC: Author. Retrieved April 28, 2008 from http://www.ncela.gwu.edu/stats/2_nation.htm

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Ortiz, A. A., & Wilkinson, C. Y. (1990). Assessment and intervention model for the bilingual exceptional student. *Teacher Education and Special Education, 14*, 35-42.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rennock-Roman, M. & Rivera, C. (2007). *Test validity and mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis.* Center for Assessment Reidy Interactive Lecture Series.

Rigney, S.L., Wiley, D.E., & Kopriva, R.J. (2008). The past as preparation: Measurement, public policy, and implications for access. In Kopriva, R.J. (Ed.) *Improving testing for English language learners* (pp. 37-63). New York: Routledge.

Rivera, C. & Collum, E. (Eds.). (2006). *A national review of state assessment policy and practice for English language learners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rivera, C., Collum, E., Willner, L., & Sia Jr., J. K. (2006). An analysis of state assessment policies addressing the accommodation of English language learners. In C. Rivera, & E. Collum (Eds.), *A national review of state assessment policy and practice for English language learners* (pp. 1-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Rivera, C., & Stansfield, C. (2001). The effects of linguistic simplification of science test items on performance of limited English proficient and monolingual English-speaking students. Paper presented at the American Educational Research Association annual meeting, Seattle, WA.

Rivera, C,.Stansfield, C. W., Scialdone, L., & Sharkey, M. (2000). *An analysis of state policies for the inclusion of accommodation of English language learners in state assessment programs during 1998-1999*. Washington, DC: Office of Bilingual Education and Minority Affairs, U.S. Department of Education.

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

Sireci, S., & Pitoniak, M. (2007). Assessment accommodations: What have we learned from research? In Laitusis, C. C., & Cook, L.L. (Eds). *Large-scale assessment accommodations: What works?* (pp. 53-66). Arlington, VA: Council for Exceptional Children.

Thurlow, M., L., Christensen, L. L., & Lail, K. E. (2008). *An analysis of accommodations issues from the standards and assessments peer review* (Technical Report 51). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 23, 2009 from http://cehd.umn.edu/nceo/OnlinePubs/Tech51/Technical51.pdf

Thurlow, M. L., Lazarus, S. S., & Christensen, L. L. (2008). Role of assessment accommodations in accountability. *Perspectives on Language and Literacy,* 17-20.

Thurlow, M. L., & Ysseldyke, J. (2002). *Including students with disabilities in assessments*. Washington, DC: National Education Association.

U.S. Department of Education. (2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001.* Washington, DC: Office of Elementary and Secondary Education.

United States Department of Education. (April 2007). 34 CFR Parts 200 and 300 Title I—Improving the Academic Achievement of the Disadvantaged; Individuals With Disabilities Education Act (IDEA); Final Rule. Washington, DC: Federal Register.

U.S. Government Accountability Office. (2007). *No Child Left Behind Act: Assistance fromeducation could help states better measure progress of students with limited English proficiency* (GAO-06-815). Washington, DC: Author. Retrieved March 1, 2009, from: www.gao.gov

Wilkinson, C. Y., & Ortiz, A. A. (1986). *Characteristics of limited English proficient learning disabled Hispanic students at initial assessment and at reevaluation*. Austin, TX: Handicapped Minority Research Institute on Language Proficiency.

Willner, L. S., Rivera, C., & Acosta, B. (2008). *Descriptive study of state assessment policies for accommodating English language learners*. Washington, DC: The George Washington University Center for Equity and Excellence in Education.

Wolf, M. K., Herman, J. L., Bachman, L. F, Bailey, A. L., & Griffin, N. (2008). Recommendations for assessing English language learners: English language proficiency measures and accommodation uses (CRESST Report 737). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved February 23, 2009 from http://www.cse.ucla.edu/products/reports/R737.pdf

Zehler, A., Fleischman, H., Hopstock, P., Pendzick, M., & Stephenson, T. (2003). *Descriptive study of services to LEP students and LEP students with disabilities: Findings on special education LEP students* (Special Topic Report 4). Washington, DC: U.S. Department of Education, Office of English Language Acquisition.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential Item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

APPENDIX A: Example Pages from the Standards and Assessment Peer Review Guidance Document

Standards and Assessments Peer Review Guidance
April 28, 2004

SECTION 4: TECHNICAL QUALITY

Critical Element	Examples of Acceptable Evidence	Examples of Incomplete Evidence
4.3 Has the State ensured that its assessment system is fair and accessible to all students, including students with disabilities and students with limited English proficiency, with respect to each of the following issues: (a) Has the State ensured that the assessments provide an appropriate variety of accommodations for students with disabilities? *and* (b) Has the State ensured that the assessments provide an appropriate variety of linguistic accommodations for students with limited English proficiency? and (c) Has the State taken steps to ensure fairness in the development of the assessments? *and* (d) Does the use of accommodations and/or alternate assessments yield meaningful scores?	The State has taken appropriate judgmental (e.g., committee review) and data-based (e.g., bias studies) steps to ensure that its assessment system is fair and accessible to all students. Review committees have included representation of identified subgroups. The State assessment system must be designed to be valid and accessible for use by the widest possible range of students. The State is conducting studies to determine the appropriateness of accommodations and the impact on test scores. **Possible Evidence** • Existing written documents describe how the principles of universal design and/or appropriate language simplification were incorporated into each of the State's assessments. • Evidence that students with disabilities were included in the test development process. • Existing written documentation of the State's policies and procedures for the selection and use of accommodations and alternate assessments, including evidence of training for educators who administer these assessments.	The State has conducted data-based bias studies but has not convened committees of stakeholders to review its assessment items. The State has convened committees of stakeholders to review its assessment items but these committees have not included representation of identified subgroups. The State assessment system is not designed to be valid and accessible for use by the widest possible range of students. The State does not have a policy on the appropriate selection and use of accommodations and alternate assessments. The State does not train or monitor personnel at the school, LEA, and State levels with regard to the appropriate selection and use of accommodations and alternate assessments. There are no appropriate accommodations for students with particular disabilities (e.g., no allowable accommodations on the regular assessment or alternate assessments for students who are visually impaired and need large print or Braille or for students who are significantly physically impaired and need assistive technology.)

List State Evidence Here

SECTION 4: TECHNICAL QUALITY

Critical Element	Examples of Acceptable Evidence	Examples of Incomplete Evidence
4.6 Has the State evaluated its use of accommodations? (a) How has the State ensured that appropriate accommodations are available to students with disabilities and that these accommodations are used in a manner that is consistent with instructional approaches for each student, as determined by a student's IEP or 504 plan? (b) How has the State determined that scores for students with disabilities that are based on accommodated administration conditions will allow for valid inferences about these students' knowledge and skills and can be combined meaningfully with scores from non-accommodated administration conditions? (c) How has the State ensured that appropriate accommodations are available to limited English proficient students and that these accommodations are used as necessary to yield accurate and reliable information about what limited English proficient students know and can do? (d) How has the State determined that scores for limited English proficiency students that are based on accommodated administration circumstances will allow for valid inferences about these students' knowledge and skills and can be combined meaningfully with scores from non-accommodated administration circumstances?	The State provides for the use of appropriate accommodations and has conducted studies to ensure that scores based on accommodated administrations can be meaningfully combined with scores based on the standard administrations. **Possible Evidence** • The State has analyzed the use of specific accommodations for different groups of students with disabilities and has provided training to support sound decisions by IEP teams. • The State routinely monitors the extent to which test accommodations are consistent with those provided during instruction. • The State has analyzed the effect of specific accommodations for students with limited English proficiency and has shared results with LEAs and schools. • Documentation of the quality and consistency of the accommodations it offers for limited English proficient students (e.g., training of translators, simplified English, standardized translation of instructions for test administration that are comparable to the regular assessment).	No analyses have been carried out to determine whether specific accommodations produce the effect intended. The State does not require that decisions about how students with disabilities will participate in the assessment system be made on an individual basis or specify that these decisions must be consistent with the routine instructional approaches as identified by each student's IEP and/or 504 plan. The State uses the same accommodations for limited English proficient students as it uses for students with disabilities.

List State Evidence Here

APPENDIX B: States' Approval Status under the Standards and Assessments Peer Review as of December 1, 2009

Name of State	Approval of Assessment System
Alabama	Approved in October 2007
Alaska	Approved by December 2006
Arizona	Approved by December 2006
Arkansas	Approved by December 2006
California	Not Yet Approved
Colorado	Approved by December 2006
Connecticut	Approved by December 2006
Delaware	Approved by December 2006
Florida	Approved in June 2007
Georgia	Approved in October 2007
Hawaii	Not Yet Approved
Idaho	Approved by December 2006
Illinois	Approved in August 2008
Indiana	Approved by December 2006
Iowa	Approved by December 2006
Kansas	Approved in May 2007
Kentucky	Approved in December 2007
Louisiana	Approved in January 2009
Maine	Approved in April 2008
Maryland	Approved by December 2006
Massachusetts	Approved in February 2007
Michigan	Approved by December 2006
Minnesota	Approved in August 2008
Mississippi	Not Yet Approved
Missouri	Approved in October 2007
Montana	Approved in November 2007
Nebraska	Not Yet Approved
Nevada	Not Yet Approved
New Hampshire	Not Yet Approved
New Jersey	Not Yet Approved
New Mexico	Approved in June 2008
New York	Approved in February 2008
North Carolina	Approved by December 2006
North Dakota	Approved in June 2007

Ohio	Approved by December 2006
Oklahoma	Approved by December 2006
Oregon	Approved in March 2009
Pennsylvania	Approved in August 2007
Rhode Island	Approved in June 2009
South Carolina	Approved by December 2006
South Dakota	Approved in October 2008
Tennessee	Approved by December 2006
Texas	Approval status unknown
Utah	Approved by December 2006
Vermont	Not Yet Approved
Virginia	Approved in December 2007
Washington	Approved in August 2008
West Virginia	Approved by December 2006
Wisconsin	Approved in December 2008
Wyoming	Not Yet Approved

APPENDIX C: Presence of ELL Issues in Critical Element Margin Comments by State

Name of State	Critical Element 1: Content Standards	Critical Element 2: Achievement Standards	Critical Element 3: Annual Assessments	Critical Element 4: Technical Quality	Critical Element 5: Alignment	Critical Element 6: Inclusion	Critical Element 7: Reporting
Alabama	No	No	No	Yes	No	Yes	Yes
Alaska	No	No	No	Yes	No	No	No
Arizona	No	No	Yes	Yes	No	Yes	No
Arkansas	Yes	Yes	No	Yes	Yes	Yes	Yes
California	No	Yes	No	Yes	No	Yes	Yes
Colorado	Yes	Yes	Yes	Yes	No	Yes	Yes
Connecticut	Yes	Yes	No	Yes	No	Yes	No
Delaware	No	Yes	Yes	Yes	No	Yes	Yes
Florida	Yes	No	No	Yes	No	Yes	No
Georgia	Yes	No	Yes	Yes	No	Yes	Yes
Hawaii	Yes	Yes	No	Yes	No	Yes	Yes
Idaho	Yes	No	No	Yes	No	No	Yes
Illinois	Yes	Yes	No	Yes	No	Yes	Yes
Indiana	No	Yes	No	Yes	No	Yes	Yes
Iowa	Yes	Yes	Yes	Yes	No	Yes	Yes
Kansas	Yes	Yes	Yes	Yes	No	Yes	Yes
Kentucky	No	No	No	Yes	No	Yes	Yes
Louisiana	No	Yes	No	Yes	No	Yes	Yes
Maine	No	No	No	Yes	No	Yes	No
Maryland	Yes	Yes	No	Yes	No	Yes	Yes
Massachusetts	No	Yes	No	Yes	No	Yes	Yes
Michigan	Yes	Yes	Yes	Yes	No	Yes	Yes
Minnesota	Yes	Yes	Yes	Yes	No	Yes	Yes
Mississippi	No	No	No	Yes	No	No	Yes

Missouri	Yes	No	No	Yes	No	No	Yes
Montana	Yes	No	No	Yes	No	Yes	No
Nebraska	Yes	Yes	Yes	Yes	No	Yes	Yes
Nevada	Yes	No	Yes	Yes	No	No	No
New Hampshire	Yes	Yes	No	Yes	No	Yes	Yes
New Jersey	No	Yes	No	Yes	No	Yes	Yes
New Mexico	No	No	Yes	Yes	No	No	Yes
New York	Yes	Yes	Yes	Yes	No	Yes	Yes
North Carolina	Yes	Yes	No	No	No	Yes	No
North Dakota	Yes	Yes	No	Yes	No	Yes	Yes
Ohio&	Yes	Yes	No	Yes	No	Yes	Yes
Oklahoma	No	No	No	Yes	No	Yes	Yes
Oregon	No	No	No	Yes	No	Yes	No
Pennsylvania	Yes	Yes	No	Yes	No	Yes	Yes
Rhode Island	Yes	Yes	No	Yes	No	Yes	Yes
South Carolina	No	Yes	No	No	No	Yes	Yes
South Dakota	Yes	Yes	No	No	No	Yes	No
Tennessee	No	Yes	No	Yes	Yes	No	Yes
Texas	Yes	Yes	Yes	Yes	No	Yes	Yes
Utah	Yes	Yes	No	Yes	No	Yes	Yes
Vermont	Yes	Yes	No	Yes	No	Yes	Yes
Virginia	Yes	Yes	No	Yes	Yes	Yes	Yes
Washington	Yes	No	No	Yes	Yes	Yes	Yes
West Virginia	Yes	Yes	Yes	Yes	No	No	No
Wisconsin	Yes	No	Yes	Yes	Yes	Yes	Yes

Note: Yes means that the state had comments relevant to ELL needs in the Critical Element. No means that the state did not have comments