Student Aims Performance in a Predominately Hispanic District

by

Lance Chebultz

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Approved April 2012 by the
Graduate Supervisory Committee:

Nicholas Appleton, Chair
Mark Duplissis
Paul Tighe

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

School districts in the United States have undergone large changes over the last decade to accommodate No Child Left Behind (NCLB). Arizona accommodated NCLB through Arizona's Instrument to Measure Standards (AIMS). Expectations were established for all students, varying by group of students based on grade, special education status, free/reduced lunch status, and English Language Learner (ELL) status. AIMS performance for subgroups has been scrutinized, due to the high stakes for schools and districts to meet expectations. This study is interested in the performance of ELL students, when compared with non-ELL students. The current study investigated AIMS performance of students in grades three through six from a large Arizona school district with predominantly low SES, Hispanic students. Approximately 90% of the students from this district were classified as ELL during their first year in the district. AIMS scores in Math and Reading were compared for ELL and non-ELL students across the years 2008, 2009, and 2010. Results suggest that there are differences in performance for ELL and non-ELL students, with ELL students scoring lower in both Math and Reading than non-ELL students. Additionally, ELL and non-ELL students showed similar performance across time in Math, with an increasing number of students Meeting or Exceeding the standards from year 2008 to 2009 for both ELL and non-ELL students. Student performance in Math for ELL and non-ELL students did not continue to improve from 2009 to 2010. On Reading performance, greater proportions of students scored as Meets or Exceeds across time for ELL students but not for non-ELL students. Non-ELL

students scored at Meets or Exceeds at equal proportions across time, although non-ELL students scored at Meets or Exceeds in higher proportions than ELL students for all three years. Results suggest the need for continued research into the appropriateness of the AIMS for ELL students and more detailed comparisons of ELL and non-ELL students within and across districts with high proportions of ELL students.

Dedication

I dedicate this dissertation to my wonderful family. Particularly to my understanding and patient former wife, Lauri, who put up with many years of research, and to our children Jordan, Jericho, and precious daughter Trinity. I must also thank my loving parents who have helped so much with baby-sitting and have given me their fullest support.

ACKNOWLEDGMENTS

I would like to thank all of those people who helped make this dissertation possible.

First, I wish to thank my advisor, Dr. Nicholas Appleton for all his guidance, encouragement, support, and patience. His sincere interest in education has been a great inspiration to me. Also, I would like to thank my committee members Dr. Paul Tighe and Dr. Mark Duplissis for their very helpful insights, comments and suggestions. Finally, I wish to thank Superintendent Dr. Raymond Aguilera for his data gathering assistance, confidence and trust in me.

TABLE OF CONTENTS

Page

CHAPTER Page

APPENDIX

LIST OF TABLES

Table Page

Table Page

CHAPTER 1

INTRODUCTION

Student AIMS Performance in a Predominantly Hispanic District

Since 1999, the Arizona Department of Education has been utilizing Arizona's Instrument to Measure Standards (AIMS) assessment as a Criterion Referenced Test for student academic skills in multiple subjects. In 2006, Arizona started requiring high school seniors to pass the three subject areas administered at the time, Reading, Math, and Writing. In 2009, the additional subject area of Science was added to the AIMS exam. Across the last decade the exam has been expanded in use to cover grades 3 through 8 and grade 10. Federal standards for performance are regulated by the No Child Left Behind (NCLB) policies. Arizona's implementation of NCLB sets goals on the AIMS exam for schools. AIMS goals are adjusted yearly to require increasingly higher percentages of students succeeding on the exam, called Adequate Yearly Progress (AYP). The expectations and performance are stratified according to subgroups (i.e. special education, English Language Learner (ELL) status, ethnicity, and Socio-economic status (SES, defined by free/reduced lunch status)).

When considering Arizona's system of requiring performance levels for subgroups, the validity of the system becomes particularly critical for schools and districts serving populations primarily falling into those subgroups. In particular, this study focused on investigating the issue of English Language Learners and their performance on the AIMS assessment. Arizona served the 6th largest number

of ELL students across the 50 states in 2007-2008, with 166,572 ELL students. Additionally, Arizona had the 3rd greatest percent of total students classified as ELL, with 15.3% of all K-12 students classified as ELL (http://www.migrationinformation.org/ellinfo/FactSheet_ELL1.pdf , 2010). Furthermore, the proportion of ELL students in a district is not the same across all districts, with many districts not having enough ELL students in subgroups to be included in calculations for AYP (minimum of 40 students per subgroup required). Districts that are located in areas of Arizona with higher populations of ELL students, such as District X at 50% ELL students, may also have confounding issues such as lower SES. The calculations for AYP in Arizona may treat districts with high populations of students in subgroups inequitably.

The current study is focused on ELL students; therefore an examination of Arizona's system for assessing and classifying students as ELL is important for understanding the criteria for a student receiving that classification. Arizona English Language Learner Assessment (AZELLA) is the augmented version of Stanford English Language Proficiency (SELP) test that was developed by Pearson and the Arizona Department of Education. This criterion referenced test is aligned to the Arizona K-12 English Language Proficiency standards. Considered a high stakes assessment, similar to AIMS, AZELLA testing materials are kept secure and protected. Five levels of materials comprise the test design in which Preliteracy is identified for kindergarten, Primary for grades 1 and 2,

Elementary encompasses grades 3, 4, and 5, with Middle Grades which includes grades 6, 7, and 8, and High School for grades 9, 10, 11, and 12.

Test design includes assessment items of multiple choice, writing samples, and short and extended oral response items. In addition to the typical assessment and response booklets for criterion referenced tests, AZELLA includes listening CD's that are mandatory for administration in Primary to High School categories in order to provide statewide consistency. With the exception of Preliteracy, all other levels contain five scoring subsets: listening, speaking, reading, writing, and writing conventions.

The current study discusses a large Arizona elementary school district referred to as District X in this document. District X is a school district serving a student population that is at an extreme in composition of SES and ELL. The district is 100% Title One, with 100% Free/Reduced lunch students. The overwhelming majority of students (greater than 90%) are Hispanic, with incoming students classified as ELL. The result of this situation is that the district struggles to comply with the NCLB regulations in a very unique environment. The ability of the district to succeed at reclassifying students into proficiency status (on the AZELLA) and for those students to succeed in the classroom; yet, continue to have difficulty with meeting the NCLB and state regulation requirements for ELL students, suggests that the NCLB and AZLEARNS requirements may be inappropriately applied to districts serving populations so high in ELL students.

District X is complying with NCLB requirements once students reach AZELLA proficiency status. Well-known and documented language barriers that exist among Structured English Immersion (SEI) students are the root causes that keep these students from reaching a Meets status on AIMS and often keep them at Approaching. (Wright, 2005) According to 2010 AIMS scores, a total of only 1,749 (51.89%) 3rd through 6th grade students were Meeting or Exceeding in reading. Another 1,111 students fell into the category of Approaches, of which 565 were identified as SEI. Districts with such a high number of ELL students have a lesser prospect of reaching Annual Measurable Objectives (AMO's) and act in accordance with NCLB requirements. Students who reclassify into Proficient status are annually replaced by non-English speaking newcomers who register during the upcoming school year. One type of newcomer student classified as an English Language Learner in this district might be considered the recent immigrant. The student may have moved to the United States in August, tested in April, and (according to NCLB) expected to assess at the Meet level on AIMS by the April date. Although exact data on the number of students that were of this type was not available, the anecdotal evidence suggests this is a non-trivial number of students.

In order to address the challenges of a student population that is primarily ELL, District X has implemented curriculum and program adjustments in an effort to meet with AYP goals. Precise implementation of Arizona's Department

of Education SEI block and SEI instructional recommendations has generated a distinguished yearly level of ELL student reclassification into proficiency status. Classrooms identified as Proficient are expected to maintain SEI best practices throughout their lessons. This is in addition to structuring the classroom in the same fashion as SEI classrooms by including intervention and small group instruction times into the schedule. District X has adopted a writing curriculum, implemented kindergarten through 8th grade that has had the effect of increasing reading scores throughout the five elementary schools. A similar pacing chart for reading instruction is being implemented for the 2010-11 school year.

Although District X has seen some success with reclassifying ELL students as proficient and meeting AYP for those reclassified students, the district's student population has presented extreme challenges in this area. One possibility for the unique challenge at District X is a potential flaw in the AYP system for evaluating district success with ELL populations. District X is very unique in its population being served, as nearly 100% of the students are classified as ELL when starting school. The methodology for determining AYP was developed using data from all districts in the state, and District X's student population is an outlier relative to other districts across Arizona. In order to address this issue, the proposed study would investigate student performance on the AIMS exam for students with different levels of ELL status. For the current study, the AIMS subject areas of Math and Reading are being examined. Specifically, the following research questions will be addressed:

1. Are there significant differences in AIMS performance (Reading and Math) for students across the levels of ELL status?
2. Does longitudinal performance on the AIMS exam differ significantly for students classified as proficient compared to students classified as ELL?

Summary

Chapter One presented an introduction to the current investigation into student performance on standardized exams in Math and Reading in an Arizona elementary school district that educates primarily ELL students. In the chapter the purpose of the study was presented, along with the specific research questions addressed in the current study. Chapter Two will review the literature in order to provide context and rationale to the current study. Chapter Three will review the specific research design used to investigate the research questions. Chapter Four will present the data analyses conducted for investigating the research questions of the current study. In Chapter Five, the results will be interpreted and discussed relative to the literature, policy, and the need for further investigation.

CHAPTER 2

LITERATURE REVIEW

Introduction

Accountability in Education

Rapid assessment has been cited as the most cost effective measure of student academic achievement compared to comprehensive school reform, increase in teacher education, computer assisted instruction, teacher salary, teacher experience, and high standards exit exams, among a host of other measures (Yeh, 2010). Cost effectiveness has been both an impetus and an objective of the call for greater accountability for better academic achievement in U.S schools (Yeh, 2010). State tests are used to provide information about the academic progress of individual students. When aggregated, the data is used to evaluate school and district performance.

In the United States, standards- and assessment-based reform was realized with the passage of two notable legislations: Goals 2000 in 1994 and the No Child Left Behind Act (NCLB) in 2001. These laws require individual students and schools to meet challenging academic standards with accountability for all students. While NCLB applies to all states, there are specific legislations, standards, and assessments developed at state levels to address the English proficiency needs of ELL students (Mahon, 2006). Following the passage of NCLB in 2001, a total of 18 states used aggregated achievement test results as the rationale for giving financial incentives to high performing and improving

schools. The same tests have also been used to sanction poor performing schools (Abrams et al., 2003). Currently, all states have aligned their education reforms with the requirements of NCLB.

In Arizona, the assessment of student learning is governed by three principal legislations: AZ LEARNS (2001), NCLB (2002), and Proposition 203 (2000). These legislations determined the development of particular tests to assess how schools meet the standards, such as the Arizona Instrument to Measure Standards (AIMS) and a system for assessing and classifying ELL students called Arizona English Language Learner Assessment (AZELLA). This literature review contains a critical review of many facets of legislation, standards, and assessments that influence the education of English Language Learners.

ELL Students in the U.S

In the United States, the number of children classified as ELL is rapidly increasing. Payan and Nettles (2008) noted that there were approximately 5.1 million English Language Learners in American schools in 2004-2005. This number translates into 10.5% of the U.S student population. An estimated 79% of all ELLs originate from Spanish-language backgrounds. While ELLs are found in every state, they are highly concentrated in Arizona, California, Texas, New York, Florida, and Illinois. These six states alone comprise 61% of the ELL student population. Additionally, the United States territory of Puerto Rico contains 1% of the total ELL population. Despite the seemingly high concentration of ELL students in the southwestern states, other states have

reported 300% or more growth in the ELL population between 1995 and 2005. These states include Alabama, Indiana, Kentucky, Nebraska, North Carolina, South Carolina, and Tennessee. California has one-third (1.6 million students) of the nation's ELL population. Approximately 85% of the ELLs in California originate from Spanish-speaking homes. Many of these ELL students in elementary and secondary learning institutions were born in the United States (http://www.educationengine.com/education_esl_factsheet.pdf).

Payan and Nettles (2008) investigated the issue of English Language Learners from both absolute numbers and from the growing number of ELL students. According to 2004-2005 data, the territories with the largest number of ELLs were Arizona; 155,789, California; 1,591,525, Florida; 299,346, Illinois; 192,764, New York; 203,283, Texas; 684,007, and Puerto Rico; 578,534. Although these states had the largest numbers of students, they were not the states with the greatest growth in ELL students. Arizona, California, Florida, Illinois, New York, and Texas did not make it onto the list of the states with the greatest growth in ELL students in the U.S., with 24 states listed. The growth was from 1994 to 2004. In the last decade, the number of ELL students has been steadily increasing in the United States, with growth appearing in many states with low numbers of students previously.

Reports of a persistent gap in academic achievement scores between native English speakers and ELLs are prevalent in educational press and popular media. Findings from 41 state agencies indicated that only 18.7% of ELL students

attained state standards for reading in English and every year the NCLB expectations rise. This leads one to consider the increasing number of ELL students as an important rationale for change in policy, especially No Child Left Behind, a re-authorization of the Elementary and Secondary Education Act (Genesee et al, 2005). The NCLB Act provides the legal framework for classifying students with limited English proficiency and lays the foundation for the operation of state legislations concerned with standard testing and accountability.

Assessing Student Learning

Historical Perspective

Standards-based assessment of student learning has become a very pervasive and influential educational topic for over a decade. The main objective of these assessments is to strengthen school and student performance and to ensure equal educational opportunities for children, particularly disadvantaged younger learners. Standards reform is supported by both Democrats and Republicans, as well as Congress and White House, the media, business stakeholders, unionists, and teachers. This high level of support is attributed to the collective desire to improve education in the United States by boosting performance scores, using better instructional methods, and ensuring that the education of American children improves relative to the global community, as evaluated by international academic assessments (Finn & Kanstoroom, 2001, p.132).

The roots of standards-based assessment of student learning can be traced to the 1983 report entitled *A Nation at Risk*. This report highlighted declining student performance according to national and international assessments. The publication of this report spurred discussions of the ideal indicators of student academic performance and ignited a shift from using school inputs, such as resources, programs, and facilities, as better indicators of educational quality. *A Nation at Risk* reiterated what had been noted in Equality of Educational Opportunity; a 1966 report by James Coleman. In response to *A Nation at Risk* and other reports describing failing public school systems, the National Commission on Excellence began pressing educators and policy makers to refocus their attention on academic outcomes. Persistent pressure led to a convention of the National Governor's Association and a summit in 1989, which is credited with setting six specific national education goals that were to be attained by the year 2000 (Finn & Kanstoroom, 2001). Of the six goals, the third goal called for students to "demonstrate competency in challenging subject matter, including English, mathematics, science, history and geography" (Finn & Kanstoroom, 2001, p. 132). The call for competency laid the framework for the development of outcome measures of performance.

The National Council of Teachers of Mathematics, the National Council of Teachers of English, the Association for the Advancement of Science, the History Standards Project, and the National Council for the Social Studies, among others, began working on subject-specific national standards. Even though initial

results provoked varying levels of controversy, these organizations supported the Goals 2000 legislation passed by Congress in 1994 and the allocation of funding to the process of developing standards. According to the legislation, the federal government would implement a compensatory education program that emphasized academic results for all students. Particularly, the legislation recognized and called on states to develop academic standards and tests to measure academic gains among disadvantaged learners (Finn & Kanstoroom, 2001).

Since the passage of the Goals 2000 legislation and NCLB, standards- and assessment-based reform has remained the premier educational strategy for measuring student performance. Basically, these legislative acts required that "all students meet challenging academic standards and that schools be held accountable for the progress of all their students" (Mahon, 2006, p. 479). These legislations promised better outcomes for ELL students.

Until NCLB, educational reform efforts had seldom recognized the needs of ELLs, despite consensus about the need for high standards and achievement expectations, the necessity of learning English as a prerequisite to accessing core curriculum and academic success, and the benefits of higher levels of English literacy and proficiency. In schools with bilingual education programs, ELL students are often given educational accommodations in which content area instruction is provided in the primary language, while English is learned as a second language (ESL) for several years before students transition to an English-

only classroom. In schools without bilingual education programs, a variety of ESL program designs are implemented to help ELL students learn and achieve academic success (Wright, 2005).

Such accommodations have in recent times conflicted with state policy restrictions on ESL and bilingual programs such as Proposition 227 (1998) in California, Proposition 203 (2000) in Arizona, and Question 2 (2002) in Massachusetts. It is important to note that these conflicts persist at a time when nationwide standardized assessments and testing are being realized under NCLB. According to NCLB, all ELL students must be included in high-stakes tests and students, teachers, and schools should be held accountable for test scores. The main rationale behind these restrictions and the inclusion of ELL students in high stake tests is that students' lack of English proficiency is a dominant factor of their overall performance and academic achievement (Wright, 2005).

In the United States, an important current issue in accountability is the poorer performance on educational outcomes by ELL students. With the rapid population growth of ELL students in Arizona, the problem is highly visible. To accommodate the linguistic and academic needs of ELL students, Arizona's educational assessment and language policies currently operate under the guidance of three federal and state language and assessment policies. These include AZ LEARNS (2001), NCLB (2002), and Proposition 203 (2000). Each policy has a specific mandate and allows specific accommodations for ELL students. AZ LEARNS is the state's high-stakes testing and school accountability

program, while Proposition 203 places restrictions on programs for ELL students (Wright, 2005). Wright observed that these policies have different mandates, so their intersection and varying interpretations reverse the accommodations they originally sought to grant to ELL students. The differing policies have led to difficulty with the implementation and interpretation of the impact, despite the intentions of the policies.

AZ LEARNS predates AIMS, the Measure of Academic Progress (MAP), and the Stanford Achievement Test (SAT-9). It was authorized in 2001 by the Arizona Revised Statutes (A.R.S.) 15-241 and uses all of the aforementioned instruments to label and provide a system through which schools can either be rewarded or sanctioned for student academic achievement (Wright, 2005).

NCLB – National System of Accountability

In a response to the proclamation that public schools are failing to provide the desired educational success for students meeting standards, the Bush administration enacted the NCLB policy in a bid to meet persistent calls for increased accountability and reliance on standardized test scores as a measure of the efficacy of education. By using standardized test scores to assess student academic achievement, NCLB sought to instill fairness and educational opportunities for all students regardless of race, ethnicity, or socioeconomic background. Research suggests that standardized test scores provide a more reliable measure of how well students are learning (Moses & Nanna, 2007). Moses and Nanna (2007) added that four major factors influenced the

implementation of standards-based reforms: administrative utility, profit motivation, political ideology, and a "testing culture" in the United States.

The NCLB Act was passed by the U.S Congress in 2001 and signed into law in January 2002 as Public Law 107-110. The legislation was a re-authorization of the Elementary and Secondary Education Act (ESEA). The purpose of the law was "to close the achievement gap with accountability, flexibility, and choice so that no child is left behind" (Preamble to Section 1). The testing and accountability sections of the highly complex law are Title I and Title III.

According to Title I, grades 3-8 must participate in annual student academic achievement testing. Title I requires all states to develop individual academic content and achievement standards and assessment tools to measure those standards, and to utilize the results of such assessments to hold individual schools accountable for student achievement. Standardized assessments must cover the domains of reading/language arts, math, and science. Every state must issue "report cards" to individual students every year. By 2014, all students should be able to pass these tests. Test score data must be disaggregated by racial and ethnic group, gender, migrant status, economic disadvantages, and limited English proficiency. Although the Act requires students who are considered Limited English Proficient (LEP) to take state tests, it also mandates states to develop valid and reliable assessments and to provide reasonable accommodations necessary for achieving English proficiency (Wright, 2005).

Title III is known as "Language Instruction for Limited English Proficient and Immigrant Students," and it commits the federal government to provide block grants to states. It requires individual states to provide block sub-grants to districts that request funding. Title III also mandates the placement of LEP students in "language instruction education programs." Transitional bilingual education and dual immersion programs are also permitted under Title III. With regard to English proficiency, Title III requires states to create ELP standards and administer proficiency assessments to LEP students annually and that schools should be held accountable for student achievement (Wright, 2005).

In compliance with the requirements of NCLB Act, the state of Arizona revamped its academic standards by developing science tests, ELL programs, and state-wide English proficiency exams. It also expanded the AIMS exams to cover not only grades 3, 5, and 8, but also grades 4, 6, and 7 (Wright, 2005).

AIMS – State/Local System of Accountability

The Arizona State Board of Education (SBE) began implementing the AIMS test to assess student achievement in 1996; however, the state-wide implementation was not achieved until 1999. The development of AIMS was achieved through a process that involved two test contractors (National Computer Systems and CTB/McGraw-Hill) in collaboration with the Arizona Department of Education (ADE), district test coordinators, and teachers. The process was comprised of selecting, writing, and revising test items to eliminate perceived bias and to ensure that items were aligned to academic content standards. A technique

called “range finding” was used to review short answer and extended writing responses before creating a uniform scoring guide (Arizona State Senate, 2008).

The AIMS tests are either criterion-referenced or non-criterion referenced tests used to assess the academic performance of students in the domains of reading, language, and mathematics (Wright, 2005). AIMS tests measure student performance and compare students in the same grades across the state to identify the performance standards for the year to comply with NCLB. Criterion referenced tests (CRT) are aligned with Arizona’s academic content standards (Arizona State Senate, 2008). The main purpose of CRT assessments is to measure student performance in core subject areas and their proficiency in line with the state’s academic standards for each content area. In Arizona, the Stanford 10 test is administered to students in grades 2 and grade 9. It covers reading, language arts and mathematics.

Currently, the AIMS test is administered in grades 3, 4, 5, 6, 7, 8, and 10 (Wright, 2005). The 10^{th} grade test essentially functions as a high school exit exam. Three forms of the test are administered. The first form, AIMS, is a dual purpose assessment test that is administered in grades 3 through 8. The criterion- and non-referenced test combines AIMS assessment items that are aligned with items on the Stanford 10 test. The second form of the AIMS test is, AIMS-HS, which is criterion-referenced and administered in 10th grade. The test is initially administered to high school sophomores, and those who do not pass may re-take the test as juniors and seniors. The test has been utilized for all students

graduating from high school since 2006. Only students who pass the test receive a high school diploma. Both AIMS and AIMS-HS assess four major subject areas: writing, reading, mathematics, and science. Grades 3 through 8 are administered Math and Reading. Grades 5 and 7 are also administered Writing. Science is only administered to students in grades 4, 8, and 10. Performance on the science section has no effect on high school graduation. The third test is AIMS-A; this test is only administered to those students with confirmed cognitive disabilities.

All non-special students, including ELLs, must take the test (Arizona State Senate, 2008). There are, however, certain exemptions for AIMS. Legislation enacted in the state of Arizona allows students with disabilities that have passed through Section 504 or an Individualized Education Program (IEP) to graduate without passing the AIMS test if these programs do not require these students to pass AIMS to graduate. Private school and home-based school pupils may also exempted from the AIMS test.

The Arizona Department of Education is responsible for determining AIMS scale scores. Currently, performance of students is classified as (1) Exceeds, (2) Meets, (3) Approaches, or (4) Falls Far Below (Arizona State Senate, 2008). With regard to the subject areas under assessment, the four classifications are determined by the state based on the scale scores. The scores are divided into ranges for each of the four classifications.

Arizona measures the number of students meeting and exceeding the acquisition of state standards based on a scale score as it appears in the table

above. Another form of accountability in the state of Arizona is AZ LEARNS, which was prompted by Proposition 301, which prohibited the use of any other language other than English during instruction. This accountability program is used by the state and by individual schools to meet the accountability requirements of NCLB. Through AZ LEARNS, the Department of Education uses individual and aggregate student results from AIMS tests to determine annual achievement and adequate yearly progress (AYP) of schools. Based on AIMS results, schools are labeled as excelling, highly performing, performing, underperforming, or failing. Test results of ELL students who have been enrolled in ELL programs for fewer than three years are not included in the calculation of academic progress (Arizona State Senate, 2008).

For those students who are Spanish-speaking only when they enter the school system, a version of the AIMS in Spanish was developed. Initially, Arizona created the Spanish-language AIMS versions for grades 3, 5, and 8. The policy required that ELL students should take it only once before switching to take the English AIMS test; however, this specification was abandoned with the passing of Proposition 203. In the Spanish version's first year of administration, there were high failure rates. Eighty-eight percent of high school sophomores and 97% of Black, Hispanic, and Native American sophomores failed at least one section of the test. Such high failure rates were attributed to difficult and ambiguous questions, reliability and validity issues with the tests, inclusion of material that was never taught in the classroom, scoring errors, and establishing

questionable passing scores. These concerns led to numerous changes in the test, as well as the pursuit of alternative tests such as the AIMS Equivalency Demonstrated (ED) tests for students who fail to pass AIMS (Wright, 2005).

Educating ELL Students in the U.S.

Historical Perspective

The movement for accountability in education expanded the implementation of standardized tests to measure student academic performance and achievement. The use of standardized tests to make high-stakes decisions regarding student placement, grade promotion, and graduation eligibility raises concerns specifically for ELLs (Solorzano, 2008). Such tests are considered "high-stakes" because their scores have the potential to directly affect a student's life options and opportunities. Test scores are used to categorize students on the basis of individual merit, as well as to allot educational placement. The original goal of utilizing scores on standardized tests was to increase equality of education for all students, but the *Nation at Risk* report and the NCLB Act of 2002 then encouraged wider use of measures of accountability and educational outcomes.

Persistent low public school performance, especially among students of color, has remained problematic for the American educational system. Latinos, who comprise the majority of ELL students, traditionally attain lower scores on standardized tests, both national (e.g., the National Assessment of Educational Progress) and in elementary, middle, and high school. Low scores suggest that students of color are inadequately prepared for and potentially ineligible to attend

college. Researchers have identified the reasons for poor performance, some of which include substandard education in schools, unprepared teachers, low academic expectations, and inappropriate program and/or instructional placement for ELL students. On the basis of these factors, standardized tests were introduced to ensure that schools are held accountable for student progress (Solorzano, 2008).

Concerns have been raised regarding clarifying the purpose for which standardized tests are developed, aligning the tests to academic curricula, and the use of and reliance on standardized test scores to make high-stakes decisions regarding allocation of financial resources and dismissal of employees. These same concerns exist regarding the addition of other language proficiency tests into established ELL instructional programs as fairness and bias are areas of controversy (Solorzano, 2008).

A student's level of proficiency is based largely on the tool used to measure language acquisition. The administration procedures of language tests, interpretation of answers, and scoring practices, all have a direct impact on the ELL status of students. In a bid to comply with accountability measures, English proficiency tests intended for use in making high-stakes decisions can have a negative effect on academic outcomes if testing policies are inappropriate, unethical, or unfair. It is for these reasons that organizations such as the American Educational Research Association (AERA) recommended that "unless a primary purpose of a test is to evaluate language proficiency, it should not be used with

students who cannot understand the instructions or the language of the test" (AERA, 2000 as cited in Solorzano, 2008). The same recommendation was voiced in the Standards for Educational and Psychological Testing, which stated that "test use with individuals who have not sufficiently acquired the language of the test may introduce construct irrelevant components to the testing process," because "for all test takers, any test that employs a language is, in part, a measure of their language skills" (Solórzano, 2008, p. 262). Therefore, concerns about the inappropriateness of standardized tests to measure academic achievement may be linked to of the underlying psychometric principles used to construct, use, and interpret standardized tests.

According to Moses and Nanna (2007), the use of high-stakes tests to measure academic achievement may degrade the aims of education, limit and constrict educational curriculum, create stress for students, constrain teachers, and reduce access to post-secondary education, thereby subverting both the learning process and the equality of educational opportunities. Some scholars have raised additional concerns regarding the inadequacy of large scale standards- and assessment-based reforms for ELL students. These reforms are "more of an aspiration than a certainty," others note that such benefits "are not a foregone conclusion," and still others state that "efforts to reform assessment as part of a systematic reform do not clearly bode well or ill for ELLs; while there are evident grounds for hope, there are no grounds for caution" (Mahon, 2006 p. 480).

Texas was the pioneer state in developing school-based testing. Starting in 1991, Texan schools began administering the Texas Assessment of Academic Schools (TAAS) test. Students who failed this state-wide test were ineligible to receive a high school diploma despite evidence demonstrating that such tests had a disproportionately negative effect on students of color. Supporters of the test noted that high school exit examinations both motivate ELL students and prepare them for college level academic work. In some cases, high school exit exams have been used to justify denying students access to a college education. Additionally, some have argued that such exams are important in bridging the gap between racial and ethnic groups as the exams motivate schools to strive for better performance, as well as individual students (Moses & Nanna, 2007).

Testing of ELLs is currently a random process to a great degree, due to inconsistent implementation and other factors that are difficult to control or uncontrollable. Testing practices and policies (for ELL students) have been grounded in deterministic views of language and linguistic groups and sometimes erroneous assumptions. The questions of "who is given tests in what language by whom, when, and where" are therefore central issues that must be considered when developing and administering an assessment system (Solano-Flores, 2008). These questions call for cultural sensitivity and relevance and appreciation of linguistic diversity (Ebe, 2010).

The Arizona English Language Learner Assessment (AZELLA)

A need was created for testing language proficiency. The impetus for the development of the Stanford English Language Proficiency test (SELP) test was the passage of the NCLB Act (2001) and Arizona's Proposition 203 (2000). The Arizona English Language Learner Assessment (AZELLA) was created using the SELP test as its model. One of the many requirements of the NCLB Act was the development of a set of English language development standards and their alignment with English language proficiency tests in every state. The Arizona Board of Education adopted SELP and began implementing it in 2004. For the next two years, the test was administered in its original form. By the fall of 2006, the test had been revised, renamed AZELLA, and implemented in the entire state. Again, this test was revised and re-released in the fall of 2009. The revised version of AZELLA is commonly referred to as Form AZ2. In this review, the test will be referred to as AZELLA. Additionally, it should be noted that Arizona is currently completing field testing of a new version of the AZELLA.

There are three main purposes of AZELLA. First, it is used to identify and classify English Language Learners for inclusion in the Structured English Immersion (SEI) program by measuring their English language proficiency (Secunda, 2009). Second, AZELLA helps to monitor the effectiveness of the SEI program through annual reassessments. This purpose satisfies the educational legislation that requires schools to abandon English language development programs that fail to yield positive results. In this way, AZELLA assures accountability. Third, the test is used in instructional planning, because it offers

instructional direction. Instructional planners can use the end goals of AZELLA to develop instruction that achieves English proficiency within the shortest time possible (http://www.mpsaz.org).

After taking the AZELLA test, students who are found to be not proficient in English get extended eligibility to participate in SEI programs. Failure to attain proficient scores on the AZELLA test also means that these students continue to be classified as ELLs. English Language Learners who score "Proficient" are then considered Reclassified as Fluent English Proficient (RFEP) and exit from SEI programs. Students continue to be monitored (assessed) for two years after reclassification to proficient status. Upon reclassification, students begin receiving English academic instruction with first English Language Learners in mainstream classrooms. Arizona legislation requires that these changes be communicated to parents; however, if a student scores less than "Proficient" in the first or second year of testing, he or she can be placed in SEI programs. The placement is not automatic; the decision to be placed in SEI programs rests with the parent, who may decide to re-enroll the student or recommend that the student receive compensatory instruction. To make an informed decision, the school must provide parents with current and accurate academic data (http://www.mpsaz.org).

Effective Education for ELL Students

The rationale behind education reform is the improvement of education for all students in the United States. Passage and implementation of NCLB, AZ LEARNS, Proposition 203, and associated programs such as AIMS and AZELLA

are major achievements of education reform in the state of Arizona. Each of these measures directly affects ELL students. Almost all these reforms are considered to be policies that place restrictions on the use of languages other than English and, therefore, neither promote nor tolerate native languages spoken by ELLs. Scholars have pointed out that the intersection of AZ LEARNS, NCLB, and Proposition 203, the three main federal and state legislations that govern standards assessment and accountability, nullifies the accommodations granted to ELL students because they are restricted oriented language policies. Nullification occurs principally due to varying, confusing, and sometimes conflicting interpretation and implementation of the policies (Wright, 2005).

In a 2003 report entitled *Betraying the College Dream*, researchers from Stanford University's Bridge Project demonstrated that confusing high-stakes tests can interfere with adequate preparedness for college. These academic assessment instruments are not aligned either to topics or to the workload in college. Again, high-stakes tests degrade the aims of education, limit and constrict educational curricula, cause increased stress to students, hold teachers accountable for reaching standardized goals regardless of the language level of their students, and reduce access to post-secondary education. As a result, the tests subvert both the learning process and the equality of educational opportunity for all students (Moses & Nanna, 2007).

Some researchers have raised concerns about over-reliance on standardized assessment tools to improve student achievement. Coleman and Goldenberg (2010) pointed out that district and school functioning should be

taken into consideration. Such aspects include school and district leadership, consistency in curricula, promotion of professional development, and the provision of continuous support and supervision in addition to regular assessment of students. States need to develop clear and uniform accountability systems that evaluate student outcomes in English language development and other subjects. According to Brooks and Thurston (2010), schools should also adopt strategies that increase ELL students' engagement in academic tasks. Brooks and Thurston demonstrated that small-group and one-on-one instruction encouraged engagement in academic tasks relative to whole-class instruction.

Huempfner (2004) faulted the one-size-fits-all approach for assessing English proficiency through standardized achievement tests for both English language and Spanish language students. According to the Huempfner (2004), such an approach is grounded in imperfect assumptions that parallel achievement tests can be used to assess bilingual and native English language students. This faulty assumption calls for more research regarding the establishment of realistic grade level English language performance standards for ELL students.

English assessment and student achievement accountability is on the top of the list of educational priorities in states with large ELL numbers. Valenzuela (2005) reiterated that, even though Texas-style accountability was the pioneer for accountability and provided the basis for the development of effective education accountability, there have been questions about how effective this accountability has been as a measure for all students, particularly Latino students. Valenzuela

posited that, although it is indisputable that schools be held accountable for student academic achievement, the high-stakes testing is inappropriate for immigrant, Latino, and English Language Learners.

The lack of ELLs meeting state standards could be improved by eliminating the requirements of Proposition 203 to increase greater flexibility in implementation of ESL and bilingual education programs. Further, not requiring ELL students to take standardized tests such as AIMS would allow educators to utilize many classroom hours spent preparing students for AIMS tests that they may fail. Not administering the AIMS assessments to ESL students would allow teachers to dedicate more time to meeting the academic, language, literacy, and cultural needs of ELLs. It will also allow schools to implement accommodations that are proven effective in mastering standards and give students enough time to learn these standards before they are required to take high-stakes exams. The NCLB Act could be repealed to eliminate fundamental flaws in labeling of schools. For instance, persistently labeling schools with a high percentage of ELL students as "Failing" denies these schools the opportunity to provide educational programs designed to enable students to become bilingual in their native language as well as English (Wright, 2005).

Contributions of this Study

Several researchers have raised concerns regarding the effectiveness and appropriateness of parallel English-Spanish academic achievement tests, particularly standardized achievement tests such as AIMS, in achieving the

educational goals of English proficiency and improved overall academic performance (Huempfner, 2004; Valenzuela, 2005; Moses & Nanna, 2007; Solorzano, 2008; Coleman & Goldenberg, 2010). For states such as Arizona that have large numbers of ELL students, the statewide implementation of AIMS means that districts and/or schools with greater numbers of ELL students coming from lower socioeconomic statuses (SES) will persistently lag behind in AYP aggregates; and, consequently, face school improvement sanctions imposed by NCLB.

This literature review presented arguments for and against the continued use of standard achievement tests in the U.S. in general and Arizona in particular. Additionally, the review discussed literature that investigated the AIMS and AZELLA as appropriate assessments for English and Spanish speaking students in Arizona. The review of literature suggests that additional research should be done to determine the appropriateness of AIMS and AZELLA tests in meeting stated education goals with regard to reading, literacy, math, and science performance among ELL students. Although there was little research on school districts with predominantly ELL students, one can assume that these issues are equally, if not more, important for such a district.

Summary

Although rapid assessment is a commonly implemented measure of student achievement because of its cost effectiveness, it also provides aggregated data that is used to evaluate the performance of the school and district as well.

Locally, it is AZ LEARNS (2001), NCLB (2002), and Proposition 203 (2000) that has cumulatively dictated the student goals, teaching guidelines, and labeled schools based on these performances. As the number of children classified as ELL continues to increase, a disparity becomes clear between the achievement scores of native English speakers and ELL students. This increasing number of ELL students in public education is a key consideration for modifying the existing policies, especially No Child Left Behind. The focus of administering these assessments is to bolster student performance and provide equal educational opportunities for disadvantaged children. As ELL students, teachers, and schools are held accountable for test scores according to NCLB, the students' lack of English proficiency is a leading factor for overall performance and academic achievement that is often not considered.

The federal government, via Title III, provided financial aid through block sub-grants and corresponding mandates that would systematically transition bilingual education and dual immersion programs out of Arizona. ELP standards were expected to be created as well as administering proficiency assessments to LEP students on an annual basis in order to be held accountable for student achievement.

There are two major assessments that affect Arizona's students: AIMS state accountability measure which is both criterion-referenced or non-criterion referenced for the main purpose of assessing student proficiency in core subjects, and AZELLA, which is used to identify and classify a child's English proficiency.

These are often considered “high-stakes” exams because their scores can potentially affect a student’s life options since scores are used to categorize students on the basis of individual merit, as well as educational placement. AZELLA results can include a student into the Structured English Immersion (SEI) program. It also assists in monitoring the level of effectiveness of the school’s SEI program through annual reassessments. These results require schools to abandon English language development programs that fail to yield positive results as indicated in legislature and this data is used in instructional planning, because it offers instructional direction.

Although commonly used, concerns have been raised that these test and their results can interfere with adequate preparedness for college since they are not aligned either to topics or to the workload in college. These tests may inadvertently degrade the aims of education, hold teachers accountable for attaining goals that are standardized and do not take language level into consideration. This can inevitably reduce a child’s access to post-secondary education and equal education opportunities.

CHAPTER 3

METHODS

Purpose of the study

This study was an investigation of student performance on the AIMS exam for students with different levels of ELL status. Student performance on the AIMS subject areas of Math and Reading was compared between levels of ELL status within one school district. Specifically, the following research questions were addressed.

1. Are there significant differences in AIMS performance (Reading and Math) for students across the levels of ELL status?
2. Does longitudinal performance on the AIMS exam differ significantly for students classified as proficient compared to students classified as ELL?

Research Design

The purpose of the study is to provide an investigation into the differences in student AIMS performance depending on students' levels on the AZELLA. An ex post facto research design will allow a comparison of AIMS performance between students classified as proficient and students classified as ELL, as determined by AZELLA. This comparison will be conducted longitudinally to investigate AIMS performance across time for ELL vs. non-ELL students. Additionally, exploratory follow-up analyses will be conducted to investigate differences identified in the omnibus statistical tests for differences.

Data Collection

The data for the current study were provided through the cooperation of a large Arizona school district (District X) serving a predominantly Hispanic population of students. District X agreed to merge necessary data files and provide the data file stripped of all identifying information. A unique identifier for each student was used in the data sets for each year in order to allow longitudinal analyses. The unique identifier was determined by the district, based on student number. Upon receipt, the data file was cleaned of records that were from out of grade range students. Additionally, records that were for other subject areas (e.g. science) were purged from the data set. After the cleaning the data, the files were loaded into a statistical software package for analyses, PASW Statistics 18, Release Version 18.0.0 (SPSS, Inc., 2009, Chicago, IL, www.spss.com). A multivariate data set was generated for conducting longitudinal analyses.

A second round of data cleaning was conducted based on preliminary demographic analyses. In the data set there were extremely low proportions of ethnicities other than Hispanic, therefore these student records were removed to provide an analysis of only the Hispanic student population. For example, the student records from 2008 included 1860 students classified as Hispanic and seven students classified as Caucasian. Due to extremely small sample sizes for the ethnic groups other than Hispanic, proper analyses would not be possible.

The data file included gender, ethnicity, free/reduced lunch status, and AIMS scores. Preliminary descriptive analyses showed that the 90% of the students were in the free/reduced lunch program. Due to the lack of variability for students in this program for the district, the free/reduced lunch variable was not investigated further in the study. The variable was found to be too highly collinear with ethnicity to provide additional insight.

Description of the District

District X is a large Arizona elementary school district serving approximately 5000 students each year. There are nine schools in the district; one preschool, six K-6th grade schools, and two 7-8th junior high schools. All the schools in the district are Title 1 with approximately 100% of the students on F/R lunch each year. Approximately 96% of the students are second language learners at the time of admission to the district, with around 30% of the students being classifies as Proficient on the AZELLA assessment each year. Schools in the district received an AZLEARNS label of Performing or Performing Plus, although only one school met AYP goals.

Description of the Sample

The student population targeted in the current study comes from a large elementary school district in Arizona, serving primarily ELL students. Students from three years (2008, 2009, and 2010) and four grades (3rd, 4th, 5th, and 6th) have been included in the study. The student needed to have a valid score on either the Math of Reading section (or both) of the AIMS exam for each year. Additionally,

the students included in the study were all Hispanic with 90% on the free/reduced lunch program. The gender was approximately 50/50 for all grades, therefore no analyses were conducted comparing gender in the current study. Demographic information for the student sample from District X meeting these requirements is presented below (see Tables 1 and 2 and 3).

Table 1

Students by year, grade, and gender from District X

	2008			2009			2010		
	F	M	Total*	F	M	Total*	F	M	Total*
3rd Grade	218	223	441	209	234	444	242	247	491
4th Grade	233	205	440	237	242	481	231	244	485
5th Grade	247	254	501	246	226	475	253	260	515
6th Grade	244	235	479	265	267	533	260	247	510

*Note: *some students were missing a gender code*

Table 2

Number of students by year, grade, and ELL Status from District X

	2008			2009			2010		
	ELL	Non	Total	ELL	Non	Total	ELL	Non	Total
3rd Grade	328	113	441	302	142	444	316	175	491
4th Grade	305	135	440	321	160	481	298	187	485
5th Grade	282	219	501	261	214	475	238	277	515
6th Grade	259	220	479	266	267	533	170	340	510

Table 3

Percentage of students by year, grade, and ELL Status from District X

	2008		2009		2010	
	ELL	Non-ELL	ELL	Non-ELL	ELL	Non-ELL
3rd Grade	74.4%	25.6%	68.0%	32.0%	64.4%	35.6%
4th Grade	69.3%	30.7%	66.7%	33.3%	61.4%	38.6%
5th Grade	56.3%	43.7%	54.9%	45.1%	46.2%	53.8%
6th Grade	54.1%	45.9%	49.9%	50.1%	33.3%	66.7%

The sample described above was used for the between-factor analyses of variance, comparing AIMS performance for ELL and non-ELL students for each year in the study. The longitudinal analyses required using a separate sample to guarantee independence of scores, one of the premises of an analysis of variance.

There were 765 students who provided data for all three years. 375 third graders and 390 fourth graders (51.9% female, 48.0% male) provided three years of data. Of those students, 71.8% were considered to have ELL status in 2008.

Data Analysis Procedures

Data analyses were completed using inferential statistical methods using a statistical software package. Between subjects analyses of variance were conducted to investigate whether or not there are differences in student AIMS performance across levels of AZELLA, for Math and Reading. These analyses are presented for each year, comparing across grade levels. Follow-up analyses were conducted, where necessary, to investigate identified overall differences.

Two-factor analyses of variance were conducted to examine student AIMS performance across time relative to the student starting level on AZELLA. The analyses used a within- by between-subject analyses of variance, for Math and Reading. The within subjects factor was AIMS performance across time for students with all three years of data (i.e. students in 3^{rd} or 4^{th} grade during 2008 and having all three years of AIMS and AZELLA data). The between subjects factor was determined by the students ELL status in 2008. As with between-subjects design above; follow-up analyses will be conducted to investigate differences.

Summary

Chapter Three presented the research design, sample, demographics, data collection, data processing, and analytic plan. The sample in the current study is

from a low-SES predominantly Hispanic population of students, with language barriers to an education. The large Arizona school district cooperated with the current study to provide data for investigation of the academic performance of students in a district with very high proportions of ELL students. The data analytic plan identifies the analyses that are presented in Chapter Four, to answer the research questions presented in this chapter.

CHAPTER 4

RESULTS

In Chapter Four, the analyses were conducted to investigate the research questions from the current study. An interpretation of the findings and a discussion of the results are covered in Chapter Five. The analyses in this chapter were organized by research question, with sections for both Math and Reading analyses under each research question. Descriptive statistics were presented, followed by inferential statistical analysis of variance tests. All statistical tests were performed with an alpha of .05. The research questions for the current study were:

1. Are there significant differences in AIMS performance (Reading and Math) for students across the levels of ELL status?
2. Does longitudinal performance on the AIMS exam differ significantly for students classified as proficient compared to students classified as ELL?

Research Question One

Are there significant differences in AIMS performance for ELL and non-ELL students?

Math Analyses

Cross-tabulations were computed to identify the number of students performing at each level of the exam for each grade and year. Students were classified into one of four categories based on performance on the AIMS exam

relative to the standards; Falls Far Below (FFB), Approaches (A), Meets (M), or Exceeds (E). The tables below show the numbers and percentages of students performing at each level on the Math AIMS exam.

Table 4

Numbers and Percentages of Students falling into each performance category on the Math AIMS exam

Year	Perf	Grade 3		Grade 4		Grade 5		Grade 6	
		N	%	N	%	N	%	N	%
2008	FFB	78	17.7	78	17.7	105	21.0	68	14.2
	A	117	26.5	92	20.9	129	25.7	86	18.0
	M	207	46.9	224	50.9	225	44.9	265	55.3
	E	38	8.6	45	10.2	42	8.4	59	12.3
2009	FFB	80	18.0	83	17.3	81	17.1	105	19.7
	A	118	26.6	103	21.4	109	22.9	106	19.9
	M	222	50.0	223	46.4	223	46.9	257	48.2
	E	23	5.2	69	14.3	60	12.6	65	12.2
2010	FFB	91	18.5	135	27.8	157	30.5	102	20.0
	A	162	33.0	137	28.2	140	27.2	117	22.9
	M	192	39.1	180	37.1	166	32.2	167	32.7
	E	44	9.0	33	6.8	52	10.1	123	24.1

The table above shows the numbers and percentage of students falling into each performance category. The percentage of students passing the exam are

greater for grade six than for the other grades, but there is variability of the percentage across grade and across year. This finding suggests that there is variability in the performance data that may be further reflected when comparing ELL and non-ELL students (the research question). In order to test for differences in ELL and non-ELL students on the exam, student performance was coded from 1 to 4, based on the level of performance on the exam (with 1 being FFB and 4 being E). A between subjects analysis of variance was computed for each grade level to determine whether or not there were significant differences in performance between ELL and non-ELL students, including grade as an additional between subject factor.

For 2008 students, the interaction of ELL status and grade level was not significant; indicating that the effect of ELL status was not dependent on grade level and the effect of grade level was not dependent on ELL status. There was a significant main effect of student grade, $F(3,1853) = 7.76$, $p < .05$. There was also a significant main effect of ELL status, $F(1,1853)=441.01$, $p < .05$. See Table 5 below for mean and standard deviations for each grade and ELL status.

Table 5

Mean Grouping for 2008 Students on Math AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	3.07	.69
	ELL	2.25	.85
Grade 4	Non-ELL	2.99	.73
	ELL	2.33	.91
Grade 5	Non-ELL	2.94	.67
	ELL	1.99	.86
Grade 6	Non-ELL	3.14	.59
	ELL	2.24	.88

For 2009 students, the interaction of ELL status and grade level was not significant; indicating that the effect of ELL status was not dependent on grade level and the effect of grade level was not dependent on ELL status. There was a significant main effect of student grade, $F(3,1925) = 3.30$, $p < .05$. There was also a significant main effect of ELL status, $F(1,1925)=447.42$, $p < .05$. See Table 6 below for mean and standard deviations for each grade and ELL status.

Table 6

Mean Grouping for 2009 Students on Math AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	2.94	.67
	ELL	2.17	.81
Grade 4	Non-ELL	3.06	.84
	ELL	2.32	.91
Grade 5	Non-ELL	3.06	.70
	ELL	2.12	.89
Grade 6	Non-ELL	3.00	.73
	ELL	2.05	.89

For 2010 students, the interaction of ELL status and grade level was not significant; indicating that the effect of ELL status was not dependent on grade level and the effect of grade level was not dependent on ELL status. There was a significant main effect of student grade, $F(3,1993) = 13.64$, $p < .05$. There was also a significant main effect of ELL status, $F(1,1993)=619.00$, $p < .05$. See Table 7 below for mean and standard deviations for each grade and ELL status.

Table 7

Mean Grouping for 2010 Students on Math AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	2.99	.69
	ELL	2.04	.82
Grade 4	Non-ELL	2.76	.80
	ELL	1.90	.85
Grade 5	Non-ELL	2.70	.89
	ELL	1.66	.80
Grade 6	Non-ELL	2.96	.95
	ELL	1.89	.92

The effect sizes for ELL status and grade have been reported in a table below. Effect sizes were considered important to include as a means of determining the importance of the variable in explaining variability in student AIMS Math scores. There are multiple effect size measures, but for this study the partial eta squared measure of effect size was used, as it represents the proportion of variability in the outcome variable that is shared with uniquely with a particular variable. The partial eta squared effect size estimates are presented for grade and ELL status below.

Table 8

Effect Sizes for Grade and ELL Status on Math AIMS Performance

	Partial Eta Squared	
	Grade	ELL Status
2008	.012	.192
2009	.005	.199
2010	.020	.237

Reading Analyses

Cross-tabulations were computed to identify the number of students performing at each level of the exam for each grade and year. Students were classified into one of four categories based on performance on the AIMS exam relative to the standards; Falls Far Below (FFB), Approaches (A), Meets (M), or Exceeds (E). The tables below show the numbers and percentages of students performing at each level on the Reading AIMS exam.

Table 9

Numbers and Percentages of Students falling into each performance category on the Reading AIMS exam

Year	Perf	Grade 3		Grade 4		Grade 5		Grade 6	
		N	%	N	%	N	%	N	%
2008	FFB	66	15.0	66	15.0	118	23.6	56	11.7
	A	160	36.3	155	35.2	175	34.9	170	35.5
	M	203	46.0	209	47.5	202	40.3	247	51.6
	E	11	2.5	9	2.0	6	1.2	5	1.0
2009	FFB	62	15.2	73	15.2	65	13.7	86	16.1
	A	162	34.5	166	34.5	143	30.1	172	32.3
	M	212	47.7	229	47.6	251	52.8	270	50.7
	E	7	1.6	9	1.9	15	3.2	5	0.9
2010	FFB	71	14.5	57	11.8	91	17.7	36	7.1
	A	155	31.6	214	44.1	173	33.6	126	24.7
	M	249	50.7	208	42.9	238	46.2	334	65.5
	E	13	2.6	6	1.2	13	2.5	13	2.5

The table above shows the numbers and percentage of students falling into each performance category. Again, the variability in the performance data suggests there may be additional variability between ELL and non-ELL students (the research question). The same coding transformation used for the Math performance data was used for the Reading data. Again, a between subjects

analysis of variance was computed for each grade level to determine whether or not there were significant differences in performance between ELL and non-ELL students, including grade as an additional between subject factor.

For 2008 students, the interaction of ELL status and grade level was significant when comparing Reading AIMS performance, $F(3,1853)=3.28$, $p < .05$. The significant interaction indicates that the effect of ELL status was dependent on grade level and the effect of grade level was dependent on ELL status. Analyses of the main effects are not meaningful when there is a significant interaction. Follow-up analyses were conducted by doing pairwise comparisons across grade for ELL and non-ELL students. To control for alpha inflation, pairwise comparisons were testing using Bonferroni and Tukey adjustments. For ELL students, grade 5 scored significantly different than grades 3, 4, and 6. Non-ELL students in grade 5 scored significantly different than grades 3 and 6, but not grade 4.

Table 10

Mean Grouping for 2008 Students on Reading AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	2.92	.55
	ELL	2.16	.74
Grade 4	Non-ELL	2.89	.54
	ELL	2.13	.74
Grade 5	Non-ELL	2.75	.55
	ELL	1.76	.70
Grade 6	Non-ELL	2.88	.39
	ELL	2.02	.69

For 2009 students, the interaction of ELL status and grade level was significant when comparing Reading AIMS performance, $F(3,1925)=2.72$, $p < .05$. The significant interaction indicates that the effect of ELL status was dependent on grade level and the effect of grade level was dependent on ELL status. Analyses of the main effects are not meaningful when there is a significant interaction. Follow-up analyses were conducted by doing pairwise comparisons across grade for ELL and non-ELL students. For ELL students, grade 6 scored significantly different than grades 3 and 4, but not 5. Non-ELL students in 2009 were at the same mean classification for all grades.

Table 11

Mean Grouping for 2009 Students on Reading AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	2.85	.54
	ELL	2.14	.73
Grade 4	Non-ELL	2.80	.64
	ELL	2.12	.75
Grade 5	Non-ELL	2.93	.49
	ELL	2.06	.75
Grade 6	Non-ELL	2.80	.48
	ELL	1.92	.73

For 2010 students, the interaction of ELL status and grade level was significant when comparing Reading AIMS performance, $F(3,1993)=2.83$, $p < .05$. The significant interaction indicates that the effect of ELL status was dependent on grade level and the effect of grade level was dependent on ELL status. Analyses of the main effects are not meaningful when there is a significant interaction. Follow-up analyses were conducted by doing pairwise comparisons across grade for ELL and non-ELL students. For ELL students, grade 5 scored significantly different than the other grades. Non-ELL students in 2010 were significantly different for grade 3 from grades 4 and 5, but not 6.

Table 12

Mean Grouping for 2010 Students on Reading AIMS by Grade and ELL Status

	ELL Status	Mean	SD
Grade 3	Non-ELL	2.97	.51
	ELL	2.09	.74
Grade 4	Non-ELL	2.79	.50
	ELL	2.05	.65
Grade 5	Non-ELL	2.77	.57
	ELL	1.83	.72
Grade 6	Non-ELL	2.89	.45
	ELL	2.12	.72

The effect sizes for ELL status and grade have been reported in a table below, as the effect sizes are important to include to determine the importance of the variable in explaining variability in student AIMS Math scores. The partial eta squared effect size estimates are presented for grade and ELL status below.

Table 13

Effect Sizes for Grade and ELL Status on Reading AIMS Performance

	Partial Eta Squared		
	Grade	ELL Status	ELL Status X Grade
2008	.027	.272	.005
2009	.007	.245	.004
2010	.020	.299	.004

Research Question Two

Does longitudinal performance on the AIMS exam differ significantly for ELL and non-ELL students?

The analyses for investigating research question two needed to be conducted with a sample that provided data points across the three years. For the current study, only students providing data for 2008, 2009, and 2010 were included. Three hundred seventy-five students from grade three and 390 students from grade four in 2008 were included in the longitudinal analyses. Gender was split 51.9% female and 48.0% male. In the sample, 71.8% were classified as ELL in 2008. Mixed model analyses of variance were conducted to investigate differences in ELL status across time. Students were collapsed into one group, removing grade level as a covariate in the analyses. Separate analyses are presented below for Math and Reading AIMS scores as the outcome variable.

Math Analyses

A between by within factor analysis of variance was conducted, with year being the within factor and ELL status the between factor. The interaction of ELL status and year was not significant. The main effect of year was significant, F (2, 762) = 9.08, $p < .05$. The effect size of year was .023. The main effect of ELL status was significant, F (1,763) = 173.35, $p < .05$. The effect size for ELL status was .185. The table below presents descriptive statistics for the average Math AIMS classification grouping.

Table 14

Mean, SD, and N for AIMS Math Grouping by Year and ELL Status

		Mean	SD	N
2008	Non-ELL	3.07	.69	216
	ELL	2.34	.86	549
2009	Non-ELL	3.21	.67	216
	ELL	2.44	.89	549
2010	Non-ELL	3.17	.81	216
	ELL	2.32	1.00	549

Reading Analyses

A between by within factor analysis of variance was conducted, with year being the within factor and ELL status the between factor. There was a significant interaction between ELL status and year on AIMS reading performance, F (2, 762) = 9.08, $p < .05$. The effect size, partial eta squared, for the interaction term

was .020, suggesting a very small effect. Main effects of year and ELL status were not interpreted, due to the significant interaction; however the effect sizes were .044 for year and .227 for ELL status. The table below presents descriptive statistics for the average Reading AIMS classification grouping.

Table 15

Mean, SD, and N for AIMS Reading Classification by Year and ELL Status

		Mean	SD	N
2008	Non-ELL	2.94	.51	216
	ELL	2.19	.72	549
2009	Non-ELL	2.94	.50	216
	ELL	2.30	.73	549
2010	Non-ELL	2.99	.43	216
	ELL	2.45	.69	549

Summary

Chapter Four presented analyses conducted to test the two research questions. Research question one (Are there significant differences in AIMS performance for students across the levels of ELL status?) was tested through analyses of variance testing for differences in AIMS performance based on grade level and ELL status. Results indicated that ELL status was the most robust factor explaining variability in AIMS performance. Research question two (Does longitudinal performance on the AIMS exam differ significantly for ELL and non-ELL students?) was investigated using between by within factor analyses of

variance testing for longitudinal changes in performance based on ELL status. Results were presented showing that ELL and non-ELL students performed significantly different from each other, but have similar changes longitudinally.

CHAPTER 5

DISCUSSION

Chapter Five focuses on the interpretation and discussion of the empirical findings for the research questions in this study. The chapter is organized in four major sections: (a) findings and interpretations, (b) limitations, (c) discussion and directions for future research, and (d) summary.

Findings and Interpretations

Chapter Four presented all the analyses conducted to answer the research questions; however, this section of Chapter Five presents the interpretation of those statistical tests. This section of Chapter Five is organized by research question and then includes an overall interpretation of the results.

Research question one asked whether or not there were significant differences in AIMS performance for ELL and non-ELL students. The question was tested by conducting a two factor analysis of variance to compare AIMS performance across grade level and ELL status. A separate analysis was conducted for each year of data included in the study (2008, 2009, and 2010). The analyses were conducted to identify whether or not there were significant interactions or significant main effects of ELL status. Either result would result in the conclusion that there were indeed significant differences in AIMS scores across ELL status. Analyses were conducted for Math performance and Reading performance separately as well. Overall this resulted in a total of six analyses of variance to provide an answer to the research question. Finally, it should be noted,

that effect sizes were examined to identify the importance of ELL status in understanding the variability of performance, as significance levels are sensitive to sample size.

The results of the analyses conducted to answer research question one were quite conclusive and consistent. For the analyses of Math performance, the results were nearly identical for each year of data. There were no significant interactions between ELL status and grade for the analyses of 2008, 2009, and 2010 data. This means that the effect of ELL status and grade are not dependent on each other, assuming there are main effects of either. Significant main effects of ELL status and grade on Math performance were found. The significant main effect of ELL status on Math Performance had an effect size between .192 and .237 for 2008, 2009, and 2010, indicating quite a large effect of ELL status on Math performance. Essentially, the effect sizes tell us that between about 20% and 24% of the variability in Math performance was attributable to ELL status. The effect sizes for grade on Math performance were between .005 and .020, indicating a very small effect of grade on Math performance. Only 1% to 2% of the variability in Math performance could be attributed to grade level.

The results of the analyses for Reading performance were somewhat different, as there was a significant interaction of ELL status and grade level on Reading performance for all three years. The results suggest that the effect of ELL status and grade depended on each other, meaning the effect of grade for ELL students was different than the effect of grade for non-ELL students. Follow-up

analyses were conducted, correcting for alpha inflation using Bonferroni and Tukey adjustments. The adjustment for alpha inflation protects from making type I errors, due to conducting a large number of statistical tests. The resulting pattern of results was examined, along with the effect sizes, and an investigation of the means. The overall interpretation is that although there were some fluctuations in Reading performance across grade that were different for ELL and non-ELL students, the ELL students always performed consistently lower than the non-ELL students. Furthermore, students in grade five tended to perform worse than students in the other grades. Grade five student performance on Reading will be discussed more thoroughly below. The effect sizes for Reading performance were similar to those for Math performance. Grade level effect sizes were between .007 and .027, indicating approximately between 1% and 3% of the variability in reading performance would be explained by grade level. The interaction effect sizes were between .004 and .005, which is small (even for an interaction). ELL status effect sizes were between .245 and .299, indicating between 25% and 30% of the variability in Reading performance was explained by ELL status.

The consistent findings for Math and Reading performance provide insight into which factors are important for students in this district. Consistency of the results across the three years included in the analyses, suggest these findings are also quite reliable. ELL status is clearly the most important factor for explaining differences in Math and Reading performance in this district, between grade level and ELL status. Although there are certainly many variables that were not

examined in the current study, the proportion of variability in Math and Reading performance that was explained by ELL status indicates that ELL status can explain a very large amount of the variability in scores for the students in District X. One issue that might help explain why grade level did not explain more variability in the Math and Reading performance is the transitioning of students from ELL to non-ELL status across time. This means that as students age into the higher grades they are increasing likely to transition out of being classified as ELL. Further investigation looking at individual student development across time might be able to provide a more thorough investigation into the relationship between ELL status and AIMS performance.

Research question two asked whether or not longitudinal performance on AIMS was significantly different for ELL and non-ELL students. This research question provides an initial investigation into student development across time, but is not looking at individual development. Rather, research question two prompted the comparison of the performance of ELL students in 2008 with non-ELL students in 2008, across three years of AIMS testing. The analyses that addressed this research question were two mixed analyses of variance. The statistical tests were two-factor analyses of variance, similar to the analyses for research question one; however, one factor was a within-subject factor and the other was a between-subject factor. A separate analysis was conducted for Math and Reading performance. The analyses were conducted on a subset of the sample included in the study for research question one. Only the students that provided data for all three years were included in the analyses for research question two.

For research question two, the results were quite similar to the results from research question one. Math performance was analyzed with year as a within-subject factor and ELL status as a between-subject factor. The interaction of year and ELL status was not significantly related to differences in Math performance. There were significant main effects for ELL status and year on Math performance, with effect sizes of .185 and .023 respectively. This means that approximately 19% of the variability in math performance across time for students was explained by ELL status, while only about 2% was explained by year. There was a significant interaction between year and ELL status on Reading performance. The interaction effect size was .02, which represents a meaningful, but small interaction. An investigation of the means revealed that there was no real change in performance across the three years for the non-ELL students, but ELL students improved across the three years. An increasing proportion of ELL students were passing the AIMS reading assessment across the three years. The ELL students performed at a lower level than non-ELL students across all three years, suggesting that they are catching up, but not very rapidly

The results of the analyses appear to provide very clear and consistent answers to the research questions. AIMS performance is significantly different for ELL and non-ELL students, for all grades and across all three years. Furthermore, ELL students and non-ELL students were significantly different in AIMS performance longitudinally. One of the most interesting differences found in the analyses was the difference between results for Math and Reading performance.

Longitudinally, Reading performance showed a very clear picture of improvements for ELL students, but no real change for non-ELL students. Across all the analyses, non-ELL students performed better than non-ELL students. This finding was represented across all grades and years. The overall results of this study show a clear difference in AIMS performance between ELL and non-ELL students, with ELL students struggling to pass the AIMS assessment at the same rate as non-ELL students.

Limitations

The current study has provided evidence of differences in ELL student performance compared to non-ELL student performance on the AIMS assessment. The study does not provide enough information to identify whether or not ELL student performance may depend, in part, on the assessments themselves and/or curriculum and ELL programs used within the district. The limitations of this study prevent drawing any conclusions about the causes for differences in performance.

The main three limitations in this study are based on the scope of the ELL students' primary language, the lack of comparative data for curriculum and ELL programs, and the aggregated research approach. The current study is constrained to only investigating Hispanic students and the curriculum and ELL programs in one school district and one state. The limit on the number of primary languages the district needed to deal with may have been beneficial for educating ELL students, as the district did not need to be concerned with multiple primary languages for students. Additionally, the curricular approach and ELL programs

could focus exclusively on transitioning Hispanic children; however there is a lack of evidence for which curriculum and ELL programs would be most successful for this type of district. The current study is limited in not being able to compare the success of the students in multiple curricula and ELL programs.

One of the most serious limitations of the current study is the aggregated approach to investigating the research questions. Although this approach was the only reasonable initial investigation of the data from a district with these characteristics, an approach that investigated the research questions at the individual student level would allow for more conclusions to be made. The current studies aggregated approach made it difficult to know how transitioning students might have adjusted means and trends in the data. The limitations of the current study prevented a thorough investigation into the research questions, but do not prevent the current study from providing an extremely sound first examination of student performance on the AIMS in an uncommon school district.

Discussion and Directions for Future Research

An investigation of student performance on AIMS reading and Math for ELL and non-ELL students in a predominantly Hispanic school district with extremely high proportions of ELL students was conducted to determine whether or not findings in the literature were supported in a district that is such an outlier.

Results from the current study suggest that there are differences in performance for ELL and non-ELL students, with ELL students scoring lower in both Math and Reading than non-ELL students. Ebe's (2010) research helps explain why the ELL scores might be lower in this population. Ebe found that students tended to do poorly on assessments when they could not relate to or bring background knowledge to the text they are asked to read. He suggested that it may not be the student's lack of knowledge or ability, but rather the texts used for the assessment. The ELL students in the current study may struggle with the cultural relevance of stories more than the students that have transitioned to proficient status. Essentially, the students in this population are almost all ELL when they arrive to the district and they begin the process of acculturation as they are learning English. Math and Reading AIMS assessments have not been developed to provide culturally relevant stories to Hispanic students, not to imply that Arizona is the only state with this issue. The multiple forms of state assessments traditionally are simple translations, not forms with culturally relevant materials. Pragmatically, the difficulties in developing culturally relevant assessments for students with every primary language are extreme for education in the United States, as there are immigrants with so many different primary languages. Ebe provides insight that might help explain the differences, but a solution to this problem presents quite a challenge for assessment in education. Development of assessments that focus on cultural neutrality may be the best solution, although this may lead to difficulties in assessing content appropriately.

Brooks and Thurston (2010) suggest that with proper teaching strategies there should be only a small gap, or no gap between ELL students and non-ELL students. In providing professional development for teachers, they advocate for increased student interaction in content area classrooms. Brooks and Thurston argue that this will promote academic engagement by ELL students. Small group and individual instruction was much more effective in maintaining student academic engagement than whole class instruction. ELL students that are given more opportunity to interact in small group and one-to-one instructional configurations showed greater academic growth than those not provided the same opportunities. The current study does not support this finding. In the current study almost 100% of the students starting school in the district were ELL. An achievement gap was present between ELL and non-ELL students in the current study with similar teaching strategies applied to all students in the district. The findings of the current study do not suggest that small group or individual instruction is not beneficial for ELL students, but rather that it may be beneficial for any students. Small group or individual instruction may help close the achievement gap, but is resource heavy. The findings of the current study suggest there may be individual differences among students for which instructional approach may be more successful. Furthermore, there is clear evidence in the current study that ELL students can transition to proficiency without small group or individual instruction. Districts, such as District X, with extremely high proportions of ELL students may be able to employ whole classroom educational

approaches that are as successful as small group or individual instruction. The benefits of small group and individual instruction may not be as pronounced in a classroom environment with almost exclusively ELL students and students that have transitioned out of ELL status and into English proficiency.

The results of this study may indicate there are areas of school and district functioning that could be changed to close the achievement gap; however, the unusual population of District X may present new considerations by educators and administrators. Coleman and Goldberg (2010) reported that ELL learners can and should be closing the academic gap with non-ELL students based on aspects of schools and districts, particularly professional development and leadership that are evidence-based with current research. They found that various aspects of school and district functioning (e.g. leadership goals, consistent curricula, professional development, ongoing support and supervision, and regular assessments that inform instruction) could bolster the academic experiences of ELL students to help close the achievement gap. Coleman and Goldberg might suggest that District X should use a uniform accountability system to measure student outcomes in academic subjects and ELL development, set high expectations, and include professional development and leadership that is evidence based. Although developing evidence-based policies in schools and districts is best practice, the current study suggests that challenges with educating ELL students may be very different for different districts. District X represents a currently unusual, but potentially increasingly common, student population,

where current research may suggest instructional approaches, curricula, and assessments that are were studied in entirely different populations and should not be generalized to this unusual district. District X may benefit from some of the suggestions of Coleman and Goldberg, but should proceed very cautiously as current research may not apply to their population of students.

The current study revealed that ELL students and non-ELL students were not improving in Reading across time at equal rates. Although the ELL students were not passing the assessment at the same levels as the non-ELL students, they were showing improvement across time. The finding that non-ELL students were performing at a higher level overall, but the proportion of students passing was not improving across time should be of some concern. This concern, as well as others raised by the current study, suggests a need for further investigation.

Future studies should consider three main issues for investigation. First, future studies should consider investigating the issue of ELL student development at the individual level of analysis. There are many factors that can influence the development of an ELL student and the transition of that student out of ELL classification is related to their level of acculturation and English acquisition, as those will determine performance on the AZELLA. Conducting future studies at the individual level, but doing longitudinal growth curve modeling, would allow the researcher to look at time varying and time invariant covariates. Second, the current study was limited to only Hispanic students in Arizona. The ELL challenge in the U.S. is not limited to Hispanic students and the curriculum

approaches and ELL programs used in Arizona can and should be compared with curricular approaches and ELL programs used in other states. Future studies should consider investigating the issue across multiple states and include variables for curriculum and ELL programs. Finally, future studies would benefit greatly from controlling for family acculturation variables. The ability of students to master the English language is largely influenced by the level of acculturation in their family and peer situation outside of school. These individual differences may be able to control for many of the differences in performance that are seen on assessments such as AZELLA and AIMS.

Summary

Chapter Five presented an interpretation and discussion of the results of analyses conducted to answer two research questions. There was conclusive evidence in the current study of significant differences in AIMS performance on Reading and Math for ELL and non-ELL students. A predominantly Hispanic and low SES school district that has a main focus of teaching ELL students English as well as traditional curriculum, ELL students Meet and Exceed expectations at a lower proportion than non-ELL students. There was also conclusive evidence that ELL and non-ELL students performed significantly different across time in Math, with both groups improving across time. For Reading, ELL students improved across time, but non-ELL students showed no significant change (although at a higher level than ELL students). The discrepancy in reading performance between ELL and non-ELL students, when taken together with the similar longitudinal

patterns in Math (although significantly different in overall performance), suggest there is a need for further investigation into ELL challenges with the AIMS assessment.

The current study offers some support for findings in the literature that criticize the current system of parallel English and Spanish academic achievement assessments. However, there is no conclusive evidence whether or not the findings are attributable to the assessments, or the educational system that has evolved around the assessments. Further research is needed to determine whether or not the assessments themselves are in need of refinement or changes in the educational curriculum and ELL programs can help close the achievement gap recognized in the literature and found in this study. In particular, districts with extremely high proportions of ELL students will struggle to meet state standards, as they stand now. The current system appears to be inequitable to such districts. With the growing population of ELL students, Arizona districts may find this to be a growing problem, as well as districts in other states.

REFERENCES

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the Classroom: Teachers Opinions of Statewide Programs, *Theory Into Practice* 42 (1): 18-29

The Arizona English Language Learner Assessment (AZELLA); http://www.mpsaz.org/elad/servesource/infoppt/files/azella_history_eld_eligib_ppt.pdf

Arizona State Senate (August 27, 2008). AIMS Test (Arizona Instrument to Measure Standards). Issue Brief, Arizona Senate Research Staff, Washington.

Brooks, K. & Thurston, L. P. (2010, Fall). English language learner academic engagement and instructional grouping configurations, *American Secondary Education* 39 (1): 45-60.Bowling Green.

Coleman, R. & Goldenberg, C. (2010, Summer). What does research say about effective practices for English learners? Part IV: Models for Schools and Districts, *Kappa Delta Pi Record* 46(4): 156-168. Indianapolis.

Daresh, C. J., & Playko, M. A. (1995). Supervision as effective staff development and inservice education. In C. J. Daresh & M. A. Playko (Eds.), Supervision as a proactive process concepts and cases (2nd ed., pp. 354-377). Prospect Heights, IL: Waveland Press.

Ebe, A. E. (2010, Sep/Oct). Cultural Relevant Texts and Reading Assessment for English Language Learners, *Reading Horizons* 50(3): 193-210. Kalamazoo.

Finn. C. E. & Kanstoroom, M. (2001). State Academic Standards, *Brookings Papers on Education Policy,* 2001, 131-179.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S Schools: An Overview of Research Findings, *Journal of Education for Students Placed at Risk* 10(4):363-385.

Huempfner, L. (2004, Fall). Can one size fit all? The imperfect assumptions of parallel achievement tests for bilingual students, *Bilingual Research Journal* 28 (3): 379-400. Philadelphia.

Mahon, E. A. (2006, Summer). High Stakes Testing and English Language Learners: Questions of Validity, *Bilingual Research Journal* 30(2): 479-488. Washington.

Moses, M. S. & Nanna, M. J. (2007). The Testing Culture and the Persistence of High Stakes Testing Reforms, *Education and Culture* 23 (1): 55-72.

Nettles, M. T. & Payan, R. M. (n.d). Current State of English-Language Learners in the U.S. K-12 student population; http://www.educationengine.com/education_esl_factsheet.pdf

Payan, R. M., & Nettles, M. T. (2008). Current state of English-Language learners in the U.S. K-12 student population. 2008 English Language Learner Symposium, Princeton, NJ. Retrieved online October 2008 from http://www.ets.org/Media/Conferences_and_Events/pdf/ELLsympsium/ELL_factsheet.pdf

Secunda, P. M. (April 20, 2009). Is Federal Court Oversight of the Arizona ELL Public School Program No Longer Necessary? Preview of United States Supreme Court Cases 36(7): 430-436. Chicago.

Solano-Flores, G. (2008, May). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners, *Educational Researcher* 37(4): 189-100. Washington.

Solorzano, R. W. (2008, June). High Stakes Testing: Issues, Implications, and Remedies for English Language Learners, *Review of Educational Research* 78 (2): 260-329.

Valenzuela, A. (2005). Leaving Children Behind: How "Texas-Style" accountability fails Latino youth, Albany: State University of New York Press.

Wright, W. E. (2005, Spring). English Language Learners Left Behind in Arizona: The Nullification of Accommodations in the Intersection of Federal and State Policies, *Bilingual Research Journal* 29(1):1-30. Washington.

Yeh, S. S. (2010, Summer). The Cost Effectiveness of 22 Approaches for Raising Student Achievement, *Journal of Education and Finance*, 36 (1): 38-75.

APPENDIX A

IRB APPROVAL LETTER

ARIZONA STATE UNIVERSITY
Knowledge Enterprise Development

Office of Research Integrity and Assurance

To:	Nicholas Appleton ED
From:	Mark Roosa, Chair Soc Beh IRB
Date:	12/20/2011
Committee Action:	**Exemption Granted**
IRB Action Date:	12/20/2011
IRB Protocol #:	1111007128
Study Title:	Investigating student performance on AIMS exams for students at different levels of ELL status

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(1) (4) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.