

Measuring Success:
Examining the Impact of Arizona's 4-Hour ELD Block.

by

Arturo Sanchez

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

Approved April 2019 by the
Graduate Supervisory Committee:

Daniel Schugurensky, Co-Chair
Eugene E. Garcia, Co-Chair
Jacob A. Chavez

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Arizona continues to implement the 4-hour English Language Development (ELD) block despite opposition and concerns regarding its appropriateness and effectiveness. Research using state and national assessments show that English language learners continue to fall behind their English proficient peers in reading achievement even after the implementation of the 4-hour ELD block. In general, there is no proven direct correlation between the program and underachievement. This study evaluated the impact of the 4-hour ELD block on reading achievement by comparing similar reading-abled students that were in the program with students that were not. The study was conducted in a district located in a predominately Hispanic poor community in order to eliminate social and economical factors that could disadvantage one group over the other. The findings demonstrated there were no significant differences in the reading achievement between both groups, supporting arguments that the 4-hour ELD block has made little to no impact for English language learners students. However, results demonstrate that early intervention may be significant in explaining increases in reading achievement.

For my father and mother.

ACKNOWLEDGEMENTS

For twenty years I have dedicated my life towards improving the educational experience for poor and minority communities. I recognize the academic barriers they face. Being a member of that community has only increased my conviction to learn and do all I can to help. I am truly grateful to all my colleagues and friends who have had the honor to share in common learning experiences and triumphs. Writing this dissertation has been an arduous process, which I would never have accomplished without their support and guidance.

I also want to recognize my committee, Dr. Daniel Schugurensky, Dr. Eugene Garcia, and Dr. Jacob A. Chavez for their help and support through my dissertation journey. I have learned volumes, both academically and practically, regarding the education of minorities and students of poverty. They have been and continue to be my mentors influencing my work in the field of education.

I want to specifically thank the Cartwright School District for allowing me to conduct this study of their schools. I appreciate their commitment towards better understanding the impact of the 4-hour English Language Development program. I am grateful to Dr. Jacob A. Chavez, the superintendent, for providing the access and resources to accomplish the task and I am most appreciative of the director, Dr. Alfredo Barrantes along with Maggie Morales who helped organize what I needed and aggregate data through relatively painless process.

TABLE OF CONTENTS

CHAPTER	Page
List of Tables	viii
List of Figures.....	ix
List of Acronyms	x
Chapter	
1 Leadership Context and Purpose of Action.....	1
Background of the Study	1
Statement of the Problem	3
Purpose of the Study.....	7
Research Questions	9
Organization of the Dissertation.....	10
A Word of Caution	11
2 Conceptual Framework and Literature Review.....	13
Conceptual Framework	13
Educating English Language Learners.....	14
Arizona and English Language Learners.....	15
Political Pressure.....	16
The 4-hour ELD block.....	16
Language acquisition theory.....	19
Structured English Immersion.....	22
Literature Review on the 4-hour ELD block.....	24
Segregation.....	25

CHAPTER	Page
Sound theory.....	26
Adequate Resources and Personnel.....	27
Proven Effectiveness.....	27
Limitations of Prior Research.....	30
Limited Research.....	30
Research Bias.....	31
High ELL Populations.....	32
Effects of Poverty.....	33
Poverty and Language Acquisition.....	34
Summary and Contribution of the Study.....	35
Conclusion of the Chapter.....	38
 3 Research Design.....	 39
Introduction to the Study.....	39
Research Design.....	39
Participants.....	41
Requirements for Participation in the Study.....	42
Quantitative Data for Analysis.....	43
Process for Identification of Participants.....	45
Statistical Analysis of Existing Data.....	48
Student T-Test.....	48
Effect Aize.....	50
AIMS Results.....	52

CHAPTER	Page
Data Gathering Process	53
Research Timeline.....	54
Adjustments to the Process.....	56
Bias and Ethical Considerations	58
Limitations and Potential Benefits of the Study.....	59
Conclusion of the Chapter	62
4 Results and Analysis.....	63
Data Set	63
Student T-Test	64
Student T-Test Analysis for Year One.	65
Student T-Test Analysis for Year Two.	65
Student T-Test Analysis for Year Three.	66
Student T-Test Analysis for Year Four.	67
Student T-Test Analysis for Year Five.....	68
Summary of Analysis	69
Effect Size	70
Effect Size Analysis for Year One.	70
Effect Size Analysis for Year Two.....	72
Effect Size Analysis For Year Three.....	73
Effect Size Analysis for Year Four.	74
Effect Size Analysis for Year Five.....	75
Summary of Analysis	77

CHAPTER	Page
AIMS Results	78
AIMS Reading Results for Year Three.	78
AIMS Reading Results for Year Four.	80
AIMS Reading Results for Year Five.	82
Conclusion of the Chapter	84
5 Conclusion And Discussion	88
Impact of the 4-hour ELD block	88
The 4-hour ELD block and meritocracy	90
Final Remarks.....	91
References	93

LIST OF TABLES

Table	Page
1. Effect Size Translated into Measured Annual Growth.....	51
2. Makeup of WPM scores for LEP and FEP Groups	63
3. List of Statistical Abbreviations	64
4. Independent T-Test Based on Spring WPM Scores for 2008 (Year 1).....	65
5. Independent T-Test Based on Spring WPM Scores for 2009 (Year 2).....	66
6. Independent T-Test Based on Spring WPM Scores for 2010 (Year 3).....	67
7. Independent T-Test Based on Spring WPM Scores for 2011 (Year 4).....	68
8. Independent T-Test Based on Spring WPM Scores for 2012 (Year 5).....	69
9. Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2007 (Year 1).....	71
10. Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2008 (Year 2).....	72
11. Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2009 (Year 3).....	74
12. Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2010 (Year 4).....	75
13. Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2011 (Year 5).....	76

LIST OF FIGURES

Figure	Page
1. Effect Size for LEP and FEP Groups.	78
2. Pass/Fail Percentages for Both Groups (Year 3).	79
3. Percent Passing (Year 3).	80
4. Pass/Fail Percentages for Both Groups (Year 4).	81
5. Percent Passing (Year 4).	82
6. Pass/Fail Percentages for Both Groups (Year 5).	83
7. Percent Passing, (Year 5).	84

LIST OF ACRONYMS

Abbreviation	Explanation
AIMS	Arizona Instrument to Measure Standards
AYP	Adequate Yearly Progress
AzELLA	Arizona English Language Learner Assessment
BICS	Basic Interpersonal Communicative Skills
BLE	Bilingual Education
CALP	Cognitive Academic Language Proficiency
EEOA	Equal Education Opportunities Act
ELD	English Language Development
ELL	English language learner
ESL	English as a Second Language
ESEA	Elementary and Secondary Education Act
ESSA	Every Student Succeeds Act
FAPE	Free and Appropriate Public Education
FEP	Fluent English Proficient
LEP	Limited English Proficient
NAEP	National Assessment of Educational Progress
NWEA	Northwest Educational Assessment
OELAS	Office of English Language Acquisition Services
PHLOTE	Primary Home Language Other Than English
R-CBM	Reading – Curriculum Based Measure
SDAIE	Specially Designed Academic Instruction in English
SEI	Structured English Immersion
SI	Sheltered Immersion
SIOP	Sheltered Immersion Observational Protocol
SY	School Year

Chapter 1: Leadership Context and Purpose of Action

Background of the Study

This research begins with “Mexican Rooms”. In my third year as a principal at G. Frank Davidson Elementary, I was interviewed on NPR to discuss the implementation of Arizona House Bill 2064 enacted in 2006 (Sanchez, 2011). The bill mandated that students not proficient in English, otherwise known as Limited English Proficient or (LEPs) would be provided English language instruction for four hours a day. This was a relatively new program with little in ways of programming information, structure, and effectiveness. The interview was focused on getting my opinion of how the implementation of the English language program, known as the 4-hour English language development block, or 4-hour ELD block for short, could potentially violate the civil rights of Hispanic students because they would be segregated for most of the school day. The language classrooms were referred to as “Mexican Rooms” because most immigrants to Arizona are predominantly of Mexican descent (Gandara, 2009). Arizona has second language learners from many countries; but due to a shared Southern border, it can be inferred that many immigrants who would be impacted by the 4-hour ELD program come from communities that spoke Spanish. In a traditional setting where most students would be Caucasian, it would be highly probable that the Hispanic students would be segregated. I did not have an answer. I had never thought of that implication because I worked at a school with students that were mostly of Hispanic descent. I would continue to ponder “Mexican Rooms” as the years passed and I was determined to find an answer.

The program was not meant to segregate but provide English language learners access to the general curriculum. The 4-hour ELD program was the state’s response to

the judge's ruling for the plaintiff in *Flores v Arizona* (1992), which held that English language learners were being removed from the regular education setting in order to participate in traditional language acquisition programs. By removing ELL students from the regular classroom, they were being excluded from the regular curriculum and therefore not receiving an equal education. The courts ruled in favor of the plaintiff and the state was ordered to provide a language acquisition program that provided equal access to the regular curriculum. The state legislature created a task force that adopted the 4-hour ELD block, a Structured English Immersion (SEI) program, which favors teaching discrete English language skills in conjunction with the regular curriculum. It would provide students who were learning English the same opportunity to advance in their academic subjects, but it would unintentionally group ELL students of the same ethnicity into segregated classrooms in areas with a high percentage of that ethnicity.

When a large majority of the student population is of Mexican origin, there will likely be Mexican rooms. G. Frank Davidson is located in an urban, metropolitan area of Phoenix known as Maryvale. The school, and the district in which it is located, has more than half its student population identified as English language learners. The majority of them speak Spanish or live in a household where Spanish is the primary language. Year after year, it was observed that a large part of the school's ELL population was identified as Limited English Proficient and were required to attend the 4-hour ELD block. In any given year, more than half of the primary grades would have students in the program essentially creating Mexican rooms. However, it was still common to have Spanish speaking minorities in regular classrooms as well. A question emerged as the early years of the program was implemented, "How different could the 4-hour ELD block be from

the regular curriculum if there was little difference in the type of students that were attending both?”

Statement of the Problem

Educational programs compete for staff, time, and money. Every year, schools determine how to spend their annual budget by allocating funds for each program. For a school leader, action research is routinely applied to determine program effectiveness in order to determine how best to use those resources. For schools whose students demonstrate academic achievement gaps, it becomes imperative to find and maintain only those programs that are most effective in reducing achievement gaps. It is required for failing schools that receive Title I grant money to support poor communities. However, determining which programs are most effective for academic achievement is not an easy task. In, “Standards Reform in High-Poverty School: The Fate of Education Reform”, Carol Barnes (2002) writes:

The problem of low achievement among poor, minority, limited-English speaking children, like many social problems, has long been under attack by educators and social scientists alike, and it has had a shifting definition. The “problem” has become more complex, and potential “remedies” perhaps more contradictory as a result of social scientists’ divergent attempts at understanding them over the years (Barnes, pg.124)

The correlation between poverty and low academic achievement brings added scrutiny for evaluating the 4-hour ELD block. Since the seminal Coleman Report released half a century ago (Coleman, 1966), many studies have shown that, due to a constellation of factors, low-income students are more likely to have lower levels of

educational achievement. Therefore, any program that is implemented in a predominately poor community must demonstrate its effectiveness. In Arizona, poor communities are often associated with a large Hispanic population; the same population that the 4-hour ELD block would impact. The larger the Hispanic population, the larger the probability more Hispanic students would be enrolled in the program and less students in the regular curriculum. It is also important to consider that it is highly probable poor minority students are misidentified for the 4-hour ELD block because Hispanics tend to score in the lowest quartile for reading and math during in the earlier grades (Gandara, 2009,).

For public schools in Arizona, schools are then required to identify English Language Learners (ELL) by proficiency level. Schools must structure their schedules to accommodate four hours of English language development for LEP students. Schools must prepare teachers to teach the 4-hour ELD block. Schools set time and resources aside for the 4-hour ELD block while also providing all students with a quality standards-based curriculum. While the program ensures that LEP students get language arts instruction at grade level, subjects like math, science, social studies, and electives have to compete for instructional time outside of the 4-hour ELD block. In schools where 50% or more of the student population qualifies for the 4-hour ELD block, the students impacted by the program and Title I requirements target the same students. Four hours dedicated to language acquisition leaves less time for those students to receive regular instruction in other subjects areas than their English proficient peers.

The 4-hour ELD block could also have an adverse affect on graduation rates. While the 4-hour ELD block is meant to support English learners, it unintentionally puts

poor minorities who may already be struggling academically at a disadvantage (Klinger, Artiles, & Mendez Barletta, 2006). It also excludes LEP students from other academic supports that English proficient peers may receive during the regular school day. The 4-hour ELD block is largely based on reading abilities. A student who is enrolled in the program exits when she or he can pass the required language exam. An exam that uses reading as a way to measure English proficiency. For Limited English Proficient (LEP) students struggling in reading, it may increase the probability of never exiting the language program until the reading deficits are addressed. This is important considering that as a student's chances of exiting an intervention decreases there is a higher risk that students develop apathy towards education adding to a school dropout increase (Kozol, 2005). Additionally, the 4-hour ELD block was designed for students to exit intervention in one year and it would be unlikely that the majority of LEP students would be English proficient in that time (Guerrero, 2004).

Strong parallels can be drawn between the 4-hour ELD block program and special education as an intervention program. Both programs focus on providing the appropriate education based on a child's need. Both look at an alternative curriculum or a modified curriculum to provide relevant grade level instruction while supporting the child's academic development through each grade until high school and beyond. The most significant difference between the two programs is how students qualify for each.

Students who are identified for special education are provided a rigorous assessment in order to qualify. Much care is taken to identify that a child as having a learning impairment that would disqualify the child from having the same learning advantages as their non-special education peers. Once identified, a child's least

restrictive environment (LRE) is carefully considered with the goal of providing the child with an academic experience most equal to their non-disabled peers. Placing a student on an alternative track for education poses risks that may negatively impact the student's chances of completing education successfully; special education staff heavily considers those implications when determining qualification. For the 4-hour ELD block, the protocols to qualify are less stringent. Identification and enrollment are based on a parent survey and a child's ability to pass a language proficiency exam. Little, if no importance, is given to a child development or academic history, the level of proficiency in the first language, or current reading ability. The result is that many more students are identified for the 4-hour ELD program than for special education.

For Title I schools, programs are evaluated primarily through assessment data. The school leader uses assessment data to identify programs that support growth as well as those programs that are not working. With this knowledge, the school leader can develop a plan toward continuous improvement by allocating more resources to effective programs while addressing concerns or terminating programs that are not as effective. Normally, school leaders have the ability to change academic programming within the limit of a district's curriculum and adopted materials. With a mandated program it is a different matter. Administrators have no choice but to implement a mandated program as instructed to do so. Changes can only be made to the law through a political process. Therefore, research, such as this study, is relevant and important in order to inform the legislators who can. Moreover, this study is timely as changes to the 4-hour ELD block are currently in discussion at the legislative level.

Therefore, The study looks to evaluate the 4-hour ELD block program to

determine how schools might best provide the language and academic needs for its English Language Learners. Information is needed to see if the 4-hour ELD block increased student achievement in reading for LEP students participating in the program. This is important to know for Title I schools where improving student achievement is the main goal of a school's purpose. Since minorities often live in poor communities, it is also important to know the impact of the 4-hour ELD block on a population already affected by a persistent achievement gap.

Purpose of the Study

The purpose of the study was to evaluate the 4-hour ELD block in order to determine the best way to provide LEP students with the language and academic instruction they need to improve academic outcomes as measured on state assessments. The study focused on a setting that has more than 50% of the student population of Hispanic decent and the primary language being Spanish. The goal was to determine how the 4-hour ELD block in regards to increasing reading achievement impacts this population. The study measured the impact of the 4-hour ELD block by comparing reading development scores of students participating in the program against similar abled students participating in regular reading classes. This was accomplished by comparing students' reading scores from the same assessments that is given to students in both groups using the same protocols. The majority of the data came from the AIMS-Web Reading – Curriculum Based Measurement® (R-CBM) while the reading scores for the Arizona's Instrument to Measure Standards (AIMS) was looked at to compare achievement outcomes. The AIMS-Web R-CBM® measures how LEP students and English proficient students progress in reading during the period of the study. The AIMS

measures how all students met the standards of the regular curriculum for grades 3-5. The study followed a cohort of LEP students and English proficient students for five school years, beginning with their first grade year.

Based on the researcher's observations as a school principal, the 4-hour ELD Block has made a positive impact in student achievement, specifically for students in the primary grades. During nine years as principal at G. Frank Davidson Elementary, the teachers generally agreed that the program was working. The staff found ways to incorporate science and social studies into the English language arts content. In time, teachers not teaching the 4-hour ELD block began using SEI strategies in their general education classrooms, believing those strategies were helpful for students struggling academically.

In the researcher's view, there is plenty of qualitative evidence to suggest that the 4-hour ELD block is effective and therefore it is justifiable to allocate a large part of school resources to it. The researcher also recognizes that little can be done to change the program and continues to adapt to the program. However, the school's achievement scores continue to demonstrate that a large portion of the student body is not meeting the standards in reading. Changes have been made to programs used in the regular curriculum. However, if most of the students not meeting the standards in reading are in the 4-hour ELD block, changes to programs in the regular curriculum will not fully address the problem. This study provides a qualitative view of the issue.

Research Questions

To address the problem identified, the study investigated the following:

1. **What is the impact of the 4-hour ELD Block in Arizona elementary schools with LEP students that make up more than 50% of the population?**
2. **What is the rate of progress made by students in the 4-hour ELD Block versus their English proficient peers in the general curriculum?**
3. **What are the possible explanations for differences in achievement between students in the 4-Hour ELD Block and their English proficient peers?**

These questions are relevant because the 4-hour ELD is an intervention that competes for educational resources such as time, staffing, training, and money. Knowing how the intervention impacts student achievement informs schools how best to allocate those resources. First, in schools with more than a 50% LEP population, there will be more ELD classrooms than regular classrooms which means a larger distribution of resources would go to the intervention than the regular curriculum. Second, Arizona has a high Hispanic population, which implicates that in schools with 50% or more LEP population, ELD classrooms would be predominately Spanish speakers; an argument held by those who see the program as promulgating segregation. Thirdly, schools with high LEP populations tend to be located in communities of poverty. The impact of poverty would affect student achievement for all students. Comparing the rate of progress between LEP and non-LEP students from a similar environment allows an analysis for

how LEP students achieve in relation to their non-LEP peers. The quantitative data from the study will assist in examining differences that exist between the 4-hour ELD block and regular instruction by supporting the qualitative observations that were made.

Organization of the Dissertation

This dissertation is organized into five chapters. The first chapter introduced the 4-hour ELD block as a language acquisition program for ELL students while receiving instruction in the regular curriculum. The chapter identified the civil rights problems that occur with incorporating a program that targets minority populations. The chapter also identified the limitation of resources when programming for the 4-hour ELD block specifically when the program and the regular curriculum targets the same students. Chapter two provides a historical background of how English language instruction has become an academic program through court rulings. The literature review focuses on how a program based on Structured English Immersion was selected above commonly practiced programs based on accepted language acquisition theories. It also introduces the effects of poverty that may have been overlooked when determining programming. Chapter three explains the research design for evaluating program effectiveness looking at the five years of data starting with the year that programming began. Chapter four applies statistical tests to compare reading achievement for LEP students and their non-ELL peers. Finally, chapter five presents conclusions and observations based on results and presents relevant points to be considered for future policy.

As a conclusion to the chapter, the point should be made that the study does not judge the 4-hour ELD block in relation to other language acquisition programs that exist. The results of the study are not meant to determine the appropriateness of the 4-hour

ELD block as a program or in a legal perspective; although they are addressed. The study simply follows that instructional leaders should determine if the 4-hour ELD block needs to be adapted, adjusted, or replaced in effort to support the demographics of the community a school serves. The study aims to apply the common evaluative tools school leaders use to measure the impact of other programs.

A Word of Caution

Caution should be given when using the terms ‘English Language Learner’ (ELL) and ‘Limited English Proficient’ (LEP). In Arizona, an LEP student is a student that is enrolled in the 4-hour ELD block based on specific requirements. An ELL student can describe any student that is learning or has learned English as a second language. This dissertation recognizes that some ELL students can and are excluded from the 4-hour ELD block because they do not meet the requirements to be enrolled. The distinction is particularly relevant when discussing ELL subgroups because ELL subgroups may have representation of students who were enrolled in the 4-hour ELD block and students that were not. For the purpose of this study, the researcher assumes that ELL students who were not enrolled in the 4-hour ELD block were enrolled in the regular curriculum along with non-ELL peers.

This study focuses on the comparison of academic achievement between LEP and non-ELL students. There are additional cautions related to comparing LEP and non-ELL students on achievement scores. First, LEP students are, by their description, “less proficient” in English and therefore it is likely that their lack of proficiency will impact academic areas that are taught and assessed in English, such as English reading. That being said, there are also varying degrees of English proficiency, from very limited to

almost proficient, that complicate the fair comparison between LEP and non-ELL students. Secondly, In Arizona, the AzELLA has been questioned regarding its reliability and/or validity as an instrument to measure proficiency (Florez, 2010). ELL students that are not LEP may not truly be proficient in English and may be subject to the same academic impediments that LEP students have.

Chapter 2: Conceptual Framework and Literature Review

Conceptual Framework

It is the intent of the literature review to create a conceptual framework of the 4-hour ELD block. Four conceptual lenses emerge from the supporting scholarship. The first two lenses focus on the background and influences that made the 4-hour ELD block the preferred English language acquisition program. The last two lenses focused on social and economic factors that impact the population being studied. Together, the case is made regarding the relevancy of the study and how it will add to the existing research.

As a language acquisition program it was important to view the program through the lens of language acquisition theory. The second lens reviews the history, legal precedent, and political policy that lead to the use of the 4-hour ELD block. This lens provides a perspective of Arizona's approach to language acquisition and its path to the 4-hour ELD block. It also provides a structure with which to evaluate the program for political purpose. The third lens reviews the current research on Structured English Immersion (SEI) in relation to the programs that also exist. The 4-hour ELD program is relatively new and much of the existing research addresses more of the program's appropriateness and less on its effectiveness. The fourth lens looks at poverty and its affect on academic achievement. Assuming that language acquisition and reading share similar developmental stages towards proficiency, the literature review looks at how poverty affects reading specifically during the early stages of development.

Educating English Language Learners. To understand how the 4-hour ELD block functions in the context of educational reform, it is important to understand how ELLs are educated in the United States. English language learners who enter the public education system have the additional burden of learning the general curriculum at the same time they are acquiring English. Prior to protections provided by the government, ELLs were immersed in the classroom with few to no supports. Students with a propensity to learn the English language quickly were more apt to succeed academically. Those students who did not were most likely to fall behind. Additionally, most ELL students are not proficient in English because they are either foreign born or of parents who are foreign born. This led to opportunities for ELLs to be discriminated against based on the origin of their race.

In 1964, Title VI of the Civil Rights Act would guarantee that students would not be discriminated against due to national origin. In 1968 the Elementary and Secondary Education Act was passed providing a free and appropriate public education for all students. The Bilingual Education Act or Title VII of 1968 would provide ELLs with a bilingual education, which focuses on supporting limited English proficient (LEP) students towards acquiring English. These protections guaranteed that ELL students would be included and supported in the public-school setting. However, it did not protect ELLs from receiving an alternative education. It wasn't until *Lau v. Nichols* (1974) that ELL students were afforded access to an equal education as their English proficient peers. With the Equal Educational Opportunity Act (EEOA) passed in congress the same year, it placed the burden on school districts to, "take appropriate action to overcome language barriers that impede equal participation".

The Every Student Succeeds Act (ESSA), the reauthorization of the Elementary and Secondary Education Act (ESEA), holds states responsible to make sure ELL students are receiving equal rights in education as their non-ELL peers. Schools are to ensure the appropriate curriculum, support, and assessments for ELL students are provided in order to access the same education as their non-ELL peers as provided under Title I. For schools struggling or failing to provide services effectively, the schools' districts were to give those schools the support needed to ensure they afford ELL students their rights under the law (The National Academy of Science, Engineering, and Medicine, 2017).

Arizona and English Language Learners. Arizona has a history of struggling to provide ELL students with a free and appropriate public education. ELL students have either been removed from regular classroom teaching to receive English language instruction or have been subject to a curriculum that is not as rigorous as that of their English proficient peers. Since most of the ELL students in Arizona are from Mexican decent, the alternative classrooms segregated minorities of Mexican decent from their Anglo peers (Gandara & Orfield, 2010). In the landmark case, *Gonzales v. Sheely* (1951), the courts found it was unconstitutional to segregate students by race. The state was forced to find a different way to teach English to their ELL student population. In the class action lawsuit, *Mendoza v Tucson School District Number 1* (1978), the courts found that ELL students of Mexican decent were unintentionally being segregated. In the case *Flores v. Arizona* (1992), the courts found that ELL students were not receiving a free and appropriate education (FAPE) as provided by the Equal Educational Opportunity Act (EEOA). In that case, the courts found that the state of Arizona was not funding

schools sufficiently to provide ELL students an equal education as their English proficient peers. The state was ordered to meet the funding requirements set by the court and was sued daily when it failed to do so (Jimenez – Castellanos et. al., 2013). This led to *Horne v. Flores* (1992) in which the state of Arizona appealed the Flores ruling. The court agreed with the state that funding was not correlated to a level of quality in instruction and reversed the fines by the original court order. However, the court did find that the state would have to ensure ELL students would receive curriculum equal to their English proficient peers. (Gandra & Orfield, 2010 and Jimenez-Castellanos et. al., 2013)

Political pressure. At the same time that the Flores cases were being resolved, there was political pressure that limited how ELLs would be taught. Arizona would pass two major pieces of legislation that would ultimately lead to the creation and implementation of the 4-hour ELD block. The first was Proposition 203, known as “English for Children” (2000), that prohibited the use of bilingual education in public schools. Prop 203 would force public schools to provide ELLs with language support and instruction through programs that relied on instruction in English only. The proposition approved language acquisition programs that preferred Structured English Immersion (SEI). The second was House Bill 2064 passed in 2006 in response to the Flores case. House Bill 2064 redefined how ELL students would be identified and provided language services as well as how those students, once in the program, could exit. It also created the English Language Task Force that would adopt the SEI language program that public schools would use. That task force adopted the 4-hour ELD block.

The 4-hour ELD block. The 4-hour ELD block that Arizona uses is a

“Structured English Immersion” (SEI) language acquisition program. It is based on the “Language Star” defined by Kevin Clark, the president of Clark Consulting and Training in California and Arizona; who also served as the consultant for the Arizona English Language Task Force. Instruction during the 4-hour ELD block focuses on five major areas of language: morphology, syntax, semantics, phonology, and lexicon (Clark, 1990). Instruction is done in English based on “English Language Proficiencies” (ELPs) or a continuum of standards (OELAS, English Language Proficiency Standards, 2019). Administrators operate under specific rules provided by the Office of English Language Acquisition Services from the Arizona Department of Education (OELAS, Monitoring, 2019).

Students are identified as candidates for the program when they are first registered for school. Parents are asked to complete a three-question survey regarding languages spoken outside of the school setting. This survey is called the Primary Home Language Other Than English (PHLOTE). The survey asks parents:

1. What is the primary language used in the home regardless of the language spoken by the student?
2. What is the language most often spoken by the student?
3. What is the language that the student first acquired?

If any of the responses of the three questions are any language other than English, the school is required to administer the Arizona English Language Learner Assessment (AzELLA) to that student. The student is then placed in the program or in the regular school program based on the results from that assessment (OELAS, 2019).

The AzELLA tests students in four areas: speaking, listening, reading, and writing. Each area has “cuts scores” to identify students into five levels of proficiency.

Students can be identified as: “Pre-Emergent”, “Emergent”, “Basic”, “Intermediate”, and “Proficient”. An overall score is also analyzed to determine which academic language band they would test into. Students who are admitted in the program are then assessed annually to measure progress towards proficiency. Students with an overall score in the “Proficient” range and are identified and labeled as “Full English Proficient” (FEP), would meet the requirement to not participate in the 4-hour ELD block program. Parents also have the right to refuse their child’s participation in the program. All students with an overall cut score in the other ranges, identified and labeled “Limited English Proficient” would be placed in a language band.

Schools are advised under the guidelines provided by Office of English Language Acquisition Services to place students at each grade level into a classroom with the same language proficiency band during the four hours of the program. Teachers are then required to use the English Language Proficiencies that best meet the needs of their students. The ELPs are structured into four ability levels with the first level being the lowest to accommodate pre-emergent and emergent language learners. Teachers are to use the ELPs to guide their instruction for each of the subject areas during the 4-hour ELD block that focus on the five major areas of language. Schools are given some flexibility in bands when there are not enough students in a band to fill a classroom or there are not enough sections in a grade level to accommodate all language bands as well as students not in the program. All adaptations must be approved by OELAS.

The 4-hour ELD block is divided into six subjects: Communication, Vocabulary, Grammar, Discrete Language Skills, Reading, and Writing. The subjects are taught within the four hours with sixty minutes dedicated each to reading and writing and thirty

minutes dedicated each to grammar, discrete language skills, vocabulary, and communication. Schools have some flexibility on how these minutes are scheduled into the day. Reading and writing must contain a full sixty minutes each. The remaining subjects could be organized in time intervals of no less than fifteen minutes each. All schedules are reviewed and approved by OELAS.

Language acquisition theory. There are different theories that explain how children in public school acquire English as a second language. The most influential theories have been by Stephen Krashen, Rod Ellis, and Jim Cummins. Their work provides the basis for the two methods most commonly used in accepted language acquisition programs. Bilingual Education (BLE) and English as a Second Language (ESL). The Arizona English Language Task Force chose a different method known as Structured English Immersion (SEI). SEI is comparatively newer in the field of education, but the components of SEI suggest it is also influenced by these accepted theories. To understand possible reasons for why the SEI method was chosen over BLE and ESL, a brief description of the major tenets is provided for each theory.

Stephen Krashen (1987) developed a theory of second language acquisition based on the belief that one would acquire a second language much in the same way that he or she learned his or her first language. He believed that the rules of a language were learned in a predictable order he called the Natural Order Hypothesis. By teaching second language learners the language rules in order, students would master the second language. Teaching the rules explicitly would also provide second language learners the opportunity to self-monitor their use of the second language. Krashen's theory hypothesized that second language learners would be motivated to learn a new language

by learning new content. Children were naturally motivated to make meaning of new content. If the content were presented in a different language than the child's first language, the child would also be motivated to learn the new language. According to Krashen, the success of learning a second language depended on the careful use of "Comprehensible Input" and in the purposeful maintenance of the learner's "Affective Filter".

Comprehensible input is a hypothesis that a second language could be acquired if the language is presented in manageable chunks. This would allow the second language learner to build from their current use of the second language. This hypothesis is similar to Lev Vygotsky's "Zone of Proximal Development". If the second language learner is presented with language structures he is not prepared for, the learning would lead to frustration and the natural tendency to be motivated by learning would be suppressed. This also holds true if the second language learner is presented language structures they have already mastered leading to boredom or apathy. The balance to keep language structures and content at a level that motivates the second language learner towards continuous learning is what Krashen called the Affective Filter Hypothesis.

Rod Ellis (1991) developed a language acquisition theory similar to Krashen's in that he believed second language learners become proficient in a second language by learning the rules of that language in a universal order he called, "the built-in syllabus". Ellis also favored the idea that a second language would be best acquired through learning content. He, however, defined learning content to be on subjects that were meaningful to the learner, not following a curriculum. Second language learners needed to learn language that would help them learn skills to survive and communicate their

needs (Ellis, 2003). He preferred that second language learners be taught “formulaic expressions” or common phrases used in the second language so that they could comprehend their environment. By focusing on teaching the language rules and phrases explicitly, a curriculum for teaching the second language could be created in a process he called “focus on form”.

Jim Cummins (1984) developed his language acquisition theory by separating language proficiency into two categories. Like Krashen and Ellis, Cummins supports the idea that second language learners are motivated to learn the second language through purposeful content. Cummins differed in that he believed students learned content to help them interact socially before learning academic content. The social structures are what Cummins called “Basic Interpersonal Communication Skills (BICS) while the more advanced academic language structures he termed “Cognitive Academic Language Proficiency (CALP). Because BICS is more social in nature and spoken in context, Cummins believes that BICS is less cognitively demanding. CALP, on the other hand, is the academic language needed to communicate and learn in a school setting. CALP would be more cognitively demanding (Cummins, 1984). Cummins also differed with Krashen and Ellis in that second language learners did not learn language rules in order. He believed that second language learners would need to be proficient in their first language before they can become proficient in the second language. The Linguistic Interdependence hypothesis postulates that a second language learner would use his or her language knowledge in the first language in order to make meaning in the second language or a process he called “Language Transfer”. According to Cummins, a second language learner would be limited in proficiency of the second language to the level he or

she is proficient in the first language; this is known as the “Threshold Hypothesis” (Cummins, 2003).

It is important to include that the National Academy of Science, Engineering, and Medicine (2017) concluded in its report that regardless of the language acquisition program chosen to teach ELL students English, ELL students benefit most when they are able to become proficient in their first language. The report emphasized that ELL students should be encouraged to orally practice their first language in the pre-school and elementary school years because oral proficiency is a crucial skill to learn how to read and comprehend at the academic level. Furthermore, schools should support ELL students towards gaining and maintaining their first language, as it will be an asset in adulthood.

Structured English immersion. Structured English Immersion (SEI) was the term written into California law when Proposition 227 was passed in 1998 and the same language was written into Arizona law in 2006. SEI methods favor second language learners being provided the regular curriculum taught in English with intensive language support to help them be proficient in a relatively short period. Kevin Clark developed SEI into a serviceable program based on the law and his research on districts that were developing programs to meet the requirements of the law. His research resulted in an instructional approach termed “English Language Development” (ELD). The ELD program focuses part of the instructional day teaching Grammar and discrete skills of the English language. The rest of the four hours would focus on Conversation, Vocabulary, Reading, Writing, that would incorporate the core curriculum modified at proficiency levels. ELL students would be able to interact with English peers during the rest of the

day through a form of “structured mixing”. This program would provide second language learners approximately four hours of language instruction built in various scheduling blocks for the proponents of the program. A generic schedule he proposed for elementary school was thirty minutes of [comprehension], one hundred twenty minutes of “Themed Literacy Instruction”, fifty minutes of Systemic ELD, and Structured Mixing as undefined (Clark, 1999 & 2009). The Language Acquisition Task Force adopted the following four-hour block schedule as follows:

- 60 minutes of conversation and vocabulary
- 60 minutes of grammar and discrete language skills
- 60 minutes of reading
- 60 minutes of writing

Based on the use of grammar and discrete skills, it could be argued that Clark favored Krashen’s Natural Order Hypothesis as well as Ellis’ “built-in syllabus”. It could also be argued that Clark follows Krashen’s hypothesis on comprehensive input and affective filters, which supports second language learners learning in English with language support. However, I found no explicit connections in my research. Clark’s writings suggested that his program model is based on what successful schools in California were doing at the time to meet California law (Clark, 1999). It is possible that those schools in California were applying language acquisition theory regardless if Clark chose to reference them or not. Clark’s writing tends to reject prior research by Cummins and Krashen because they supported teaching LEPs in their primary language (Clark, 2009). With limited research to go on, it seems that Clark’s 4-hour ELD block gained prominence because it emphasized teaching LEPs in English, an approach voted into law by California and Arizona.

It is important to note that at the same time of Clark's writing there were two other methods with similar structures being used in California, which may have influenced Clark's work. The first one was "Sheltered Immersion" (SI), which in California was referred to as the "Specially Designed Academic Instruction in English (SDAIE) (Genesee, 1999). The second was the "Sheltered Immersion Observation Protocol" (SIOP) (Echevarria, Vogt, & Short, 2000). Both incorporated some of the same approaches in regards to teaching grammar and incorporating core curriculum with language instruction. In a 7-year study of the SIOP model, there was evidence that teachers using the SIOP model had students do "slightly better" than a comparative group on a writing task (Echevarria, Short, & Powers, 2006). Since these methods were not written into law, I chose not to research them further. However, there may be anecdotal evidence that school districts incorporated them into the 4-hour ELD block adding a factor to program effectiveness. This could be a topic for a future study.

Literature Review on the 4-hour ELD block

During my early research for this study, little was written on the impact of the 4-hour ELD block in regards to academic achievement. There are volumes written about language acquisition, English as a Second Language, and Bilingual Education. This was to be expected since SEI was relatively new to the education setting and Kevin Clark's program became relevant in California and Arizona where "English Only" laws were put in place. Early articles were more focused on whether the program was appropriate. For instance, Karen Lillie (2010) examined the potential issues the 4-hour ELD block failed to address. Issues consistent with the three-pronged tests defined in *Castañeda v. Pickard*

(1981). In that case, it was found that a language acquisition program was deemed to be appropriate if:

1. The program had to be from sound educational theory.
2. The program would be implemented with adequate resources and personnel.
3. The program demonstrated to be effective in overcoming language barriers over time.

Too little time has passed for research to determine if the program proved effective. After ten years, the research on program evaluation is emerging. However, it seems that research is slow in coming, possibly since SEI is only implemented in a few states. Continued research on program evaluation of the 4-hour ELD block is needed to influence state policy. Fortunately or unfortunately, SEI doesn't seem to be gaining popularity. However, it does affect education policy in the state of Arizona, creating the incentive to study it.

Segregation. One concern regarding the 4-Hour ELD block that is perceived to take a large part of the debate is segregation. The 4-Hour ELD block requires LEP students to be placed in language bands at each grade level. Although one could argue that students would only be required to remain together for four hours of the day, in elementary schools, switching classrooms for math, science, and exploratory classes is not often practiced. The practice of elementary schools keeping classes together would mean students would remain together for the whole day. If these schools would be located in communities with high populations of a common language, such as Spanish, then it would be very likely that it would inadvertently place many students of the same race or ethnicity into one group. For Arizona, which has a high population of people of

Mexican descent, the odds favor potential “Mexican Rooms” (Gandara. & Orfield, 2010). Therefore, it is possible high ELL populations may affect the outcomes for language acquisition instruction. Districts with higher ELL populations may have considerably different outcomes affected by how students interact with each other. In general, the students in schools with a higher ELL population may not be exposed to or practice sufficiently common English words and phrases that would accelerate individual student’s proficiency in English (Slavin & Cheung, 2005).

Sound theory. Opponents of the 4-hour ELD block argue that there are other language acquisition programs that have a stronger scientific basis than Structured English Immersion, preferring bilingual programs to most. It seems fewer researchers support the Structured English Immersion approach, which favors the learning of English in one year through the 4-Hour ELD block model (Guerrero, 2004). “For many years, researchers, educators, and policymakers have debated the question of the appropriate language of reading instruction for children who speak language other than English.” One study sought to resolve the debate regarding what was the appropriate language for ELL students to receive reading instruction. The study used a “best-evidence analysis” of different studies on Bilingual Education and Structured English Immersion. The meta-analysis reviewed published studies of both bilingual education and Structured English Immersion methods to determine which of the two were considered more effective based on the results of each type of study. The study suggested that non-English proficient students need to have some form of instruction in their first language in order acquire English well; specifically, to access content knowledge taught in English. Based on the analysis, the study drew the conclusion that bilingual programs yielded better results than

SEI (Slavin & Cheung, 2005). It was also argued that the state failed to identify the 4-hour ELD block as “significantly more effective” than existing programs (Martinez-Wenzel, Perez, & Gandara, 2010) or even providing little research as to its effectiveness (Martinez-Wenzel et. al., 2010).

Adequate resources and personnel. It was also argued that implementing the 4-hour ELD block would create a strain on school resources, which would negatively affect graduation rates for those in the program. (Karen E. Lillie et. al., and Peer & Perez, 2010). A common perception was that putting students in proficiency bands would require more teachers per grade level. However, I did not find any study that established a correlation between hiring more teachers and the rate of graduation for LEP students. One study argued that policies were poorly planned when taking account financial support and teacher training (Rios-Aguilar, Gonzalez-Canche, & Moll, 2010). It suggested that teachers would not be efficiently trained in SEI methods to support LEP students. A few years later, the same researchers conducted a qualitative study of Arizona teachers analyzing their perceptions of the 4-hour ELD block (Rios-Aguilar, Gonzalez-Canche, & Moll, 2012). That study suggested that teachers who provided SEI services held a more negative view of the program. It has also been argued that the AzELLA may not be a valid test for determining English proficiency because the AzELLA cut scores used to identify the levels of proficiency have not been sufficiently normed (Florez, 2012). This may be a factor to understand how students are identified and exited, but it is not the focus of this research.

Proven effectiveness. The National Assessment of Education Progress (NAEP)

found that ELL Hispanic students performed significantly lower than non-ELL Hispanic peers in Reading for the test given in 2009. The 2009 test showed a 29-point gap between ELL students and their non-ELL peers for 4th grade and a 39-Point gap between the same sub-groups in 8th grade. It also reported that the achievement gaps on the NAEP given in 1998 were not significantly different between the same sub-groups for the same grades (NAEP, 2011). Although there may be many factors to consider, the data demonstrate that the gap has increased significantly since 1998.

According to research in the field, the gap between FEP students and their English proficient peers continues to grow in Arizona since the 4-hour ELD Block was implemented. For instance, Jimenez- Castellanos, Combs, Martinez, & Gomez (2013) reported that,

AIMS standardized test results for the 2002-03 school year from 3rd grade to high school indicate that only 3rd grade math exceeded the annual measurable objective goals set for English Language Learners. Comparatively, in 2009-10, none of the ELLs from 3rd grade through high school met the passing annual objective goals. In fact, the higher the grade, the lower the percentage of ELLs who passed the assessment (Jimenez-Castellanos, et. al., pg.13).

AIMS data results from 2005 through 2009 reported that there was only a 2.6% increase in reading for ELL students suggesting that achievement gaps between English Language Learners and non-ELL students persists despite the implementation of the 4-hour ELD block (Garcia, Lawton, & Diniz de Figuerido, 2010). The report states, “... these policies have generated no substantive decrease in achievement gaps and, in comparison to other states without such restrictive policies, Arizona’s achievement gaps

are significantly greater.” Garcia, Lawton, & Diniz de Figuerido (2012) published a second paper that provided more specific measures for the same study. The data showed that the rate of improvement for ELL and non-ELL subgroups was similar for each grade level for 2005-2009. However, the achievement gap was virtually consistent with ELL subgroups achieving significantly lower than their non-ELL peers for each year of the study. The same report included a comparative study between Arizona schools and those in Utah and Washington DC. Utah and Washington were chosen for the study because; “... these educational entities do not restrict [programmatic] arrangements, legally, as does Arizona.” The findings reported that ELL students in Arizona were consistently achieving lower than DC schools and much lower than schools in Utah (Garcia, Lawton, & Diniz de Figuerido, 2012).

In a study that measured AIMS and Terra Nova scores between 2006 and 2010 for students in grades two through twelve (C-Rios Aguilar, Gonzalez Canche, & Sabetghdam, 2012), two tests allowed the research to follow grade two with the Terra Nova and grades three through twelve with AIMS. Their findings seem to support the hypothesis that there is no significant difference in AIMS scores between ELLs and non-ELLs for reading. It did find that ELLs who were receiving language development support but were not in the 4-hour ELD block yielded larger magnitude in positive results but were still not significantly different. The researchers concluded the 4-Hour ELD block did not increase academic achievement scores for ELL students. The findings are significant if proven that ELL students tend to do better in regular classroom. It would mean that students in schools with high ELL populations and attend the 4-hour ELD block would be at a disadvantage.

Based on research using NAEP and AIMS data a strong argument can be made that since the 4-hour ELD block was implemented, students in the ELL subgroup may have made some improvement in reading but it is still significantly lower than the improvements for their non-LEP peers. According to the research, the 4-hour ELD block has not been effective towards closing achievement gaps. The data from the California districts analyzed by Kevin Clark, in his studies suggested that the program had a positive impact in closing the achievement gap (Clark, 1999). However, Clark is the creator and founder of the program, which suggests a possible bias leading to unreliable research. Another study of the Structured Immersion Observational Protocol (SIOP) model, which is similar to the 4-hour ELD block, demonstrated achievement growth for the ELL students in the 7-year study (Echevarria, Powers, & Short, 2006). Both studies demonstrated academic growth was found at the school level.

A pattern emerges when examining the research on the impact of instruction for ELL students. Based on the research, when achievement scores are measured at the school level, ELL students tend to demonstrate growth indicating that the SEI program is being effective. It is implied, however, that when the students are included in the ELL subgroup at the state level, that achievement growth is considered insignificant. There is not clear indication as to why the pattern exists.

Limitations of Prior Research

Limited studies. A review of the literature regarding the Structured English Immersion (SEI) model, more specifically the 4-hour ELD block, is limited in relation to

other language acquisition programs based on Bilingual Education (BLE) and English as a Second Language (ESL). It may be due to SEI being a relatively new program in education. It can also be argued that the SEI model does not seem to be widely practiced in U.S. schools. Interestingly enough, the states adopting the SEI or SIOP models are those that have adopted an “English Only” approach to classroom instruction. Unlike BLE and ESL, SEI and SIOP models are the only models that restrict teaching in any other language other than English. The majority of the research on the SEI model and related programming, either for or against its appropriateness, tend to come from California and Arizona; the two states where the model is practiced statewide. With limited examples of the SEI model being practiced in public schools at the national level and with a short time period for implementation, it would be expected that little research have been done on the topic. However, working in a state that has mandated the program, it continues to be important that research be conducted on the SEI model – and specifically the 4-hour ELD block - as it makes a large impact on how schools provide educational services and policy its student populations.

Research bias. Taken as whole, what little research is given on the SEI model and programs based on the model, there tends to be a bias towards the appropriateness of the program. Current research tends to favor language acquisition programs based on more traditional models, which incorporate some of the ELL’s first language. It follows that there is more supporting data on bilingual based language acquisition programs, which have been in practice longer, than programs based on SEI. This alone suggests there would be more resources for schools to choose from, established organization models for implementing the program, and research-based training programs for teachers.

The research makes a compelling argument that the SEI model was mostly likely chosen for its basis on English instruction, which met the existing law for the state. However, a counter argument can be made that ELL students were failing when bilingual-based language acquisition models were in place.

For professionals in the states that must implement the SEI or SIOP models by law, this is all rather a moot point. The SEI model and the 4-hour ELD block program have met two of the requirements established in the Pickard ruling. The courts have established that the SEI model is based on sound language acquisition theory. The state has also met the requirements, identified by the courts, of resourcing the program adequately. The final requirement of the Pickard ruling consists of data demonstrating that the program is effective in overcoming language barriers over time. It is this final requirement that still needs time for enough data to be acquired. This study proposes to add to that body of research.

High ELL populations. A report by the Pew Hispanic Center indicated that high populations of ELL students in a school tend to negatively affect academic achievement.

In all five states investigated [Arizona being one of them] and irrespective of grade levels ELL students were much less likely than white students to score at or above the states proficiency level. However, when ELL students attended public schools with a minimum threshold number of white students the gap between the math proficiency scores of white students and ELL students was considerably lower, The Pew Hispanic Center analysis has found (Fry, pg.1).

While this adds to the concerns regarding segregation as a violation of civil rights, it also suggests that students in the 4-hour ELD block may face social factors that limit

the ability to acquire English and hence further inhibit academic achievement, particularly, in communities that have high ELL populations. Under these conditions, it is important to ask if there is any substantial difference for students participating in the 4-Hour ELD block from students who are not.

To complicate matters, ELL subgroups in state and national tests do not make a distinction between English language learners (ELL) and students who are limited English proficient (LEP). According to the law, parents who register students in an Arizona school must participate in the Primary Home Language Other Than English (PHLOTE). If a parent reports that their child has experienced a language other than English, that student can be considered an ELL. Those students must take the language proficiency exam AzELLA. Based on the results, they will either be identified as LEP and be enrolled in the 4 – hour ELD block program or as Full English Proficient (FEP) and be enrolled in the regular curriculum. Regardless of the language proficiency identification, the student is still considered as ELL. Communities with high ELL populations would tend to have more instances of ELL students participating in both the 4-hour ELD block and the regular curriculum.

Effects of poverty. A factor that is not often included in the SEI research is the effect of poverty on student achievement. Schools located in areas of poverty tend to have higher achievement gaps than those in more affluent communities, and students of minority groups tend to be in schools of poverty (Coleman, 1966). The social and economic constraints of poverty affect all children of poverty regardless of race or ethnicity (Burney & Bielke, 2008). The research on the effects of poverty are relevant to this study considering that there were 6.1 million Latino children living in poverty in

2010 making up 37.3% of the children population, according to a Pew Hispanic Center report (2010). This is significant because that year Latino children made up only 23.1% of the U.S. children population for the same year (Pew Hispanic Center, 2011). The census data indicates that a Hispanic child is more likely to be living in poverty and attending a school that receives Title I funds.

Poverty and language acquisition. Poverty does not only affect student achievement but also affects language development. Hart & Risley (2003) conducted a study of families from three socio-economic levels examining language development for children ages 0 – 3. They found that children of poverty were less likely prepared to learn how to read by kindergarten due to a deficiency in language experiences. They found that children of poverty did not engage in robust communication with caregivers between the ages of 0 and 3. The study estimated that children of affluent families engaged in 30 million more words than their peers living in poverty. This lack of communication created a deficit in foundational skills for “reading readiness” (Hart & Risley, 2003). Basil Bernstein’s (1971) work on restricted codes and elaborated codes supports the Hart & Risley conclusion. He explained that restricted codes are words and phrases spoken within a family or community that do not follow formal grammatical or syntactical rules of the language being used. The words or phrases that are used by the immediate community are only understood among that community. The elaborated codes would be the use of language governed by grammar and syntactical rules and used by the larger community, such as a formal school setting (Bernstein, 1971).

Making the assumption that Hispanic ELL students are more likely to live in poverty, It is argued that Hispanic ELL students are lacking in strong communication

exercises and predominantly speaking in restricted codes before starting school thereby creating a disadvantage for “reading readiness” in both their first and second language. If Hispanic ELL students are not proficient in their first language it is more likely that they will not become proficient in their second language, a condition known as semilingualism (MacSwan, 2000). There is also research that suggests ELL students who do not become proficient in phonemic awareness in their first language in kindergarten will demonstrate difficulties with reading later (Atwill et. al., 2010). Subsequently, the Threshold Hypothesis (Cummins, 2003) suggests that if Hispanic ELL students develop a reliance on restricted codes in their first language, they increase the risk of struggling to learn elongated codes in their second language – English – at school. What all this implies is that Hispanic ELL students that live in poor communities are more likely to not be proficient in both their first and second language when they start school. It also implies that Hispanic ELL students that live in poor communities are experiencing the same disadvantages as the non-ELL students living in the same communities.

Summary and Contribution of the Study

The literature review overwhelmingly supports that the SEI model, which includes the 4-hour ELD block, has not been proven to be the most appropriate program for ELL students. The literature reflects that bilingual-based language acquisition programs have demonstrated, through extensive research and studies, to be superior than the SEI model. It also concludes that the SEI model is not fully planned out to support school on how to implement it effectively. The literature also reflects that the SEI model's close association with “English Only” laws is discriminatory and promotes segregation. Although the literature makes a strong case that the SEI model is not the

most appropriate, there is little in which to determine if the SEI model is effective.

The current research on reading achievement for ELL has demonstrated that ELL students made gains in reading since the 4-hour ELD block was implemented. However, reading achievement scores continue to be lower than non-ELL groups. Studies of the 4-hour ELD block conclude that achievement results in reading for students in the program are not significantly different than results of students not in the program. It implies then that the 4-hour ELD is increasing reading achievement but not at the rate needed to close the achievement gap. Additionally, it suggests that the 4-hour ELD block is no more effective than the regular reading curriculum already in place. However, the review of literature has limitations in the research that may provide a more conclusive evaluation of the program.

The first limitation of the research is assuming that all ELL students are participating in the 4-hour ELD block. The research relies on state and national assessments in reading that group all types of ELL students as one for the purpose of evaluation. There is a large range in abilities for ELL students including those who are proficient in English as well as those enrolled in special education. Furthermore, there are many ethnicities beyond Hispanics that represent the ELL subgroup as there are varying social-economic classes. Not all ELL students are in fact enrolled in the 4-hour ELD block; only those identified as LEP are. Therefore, it is difficult to correlate the reading achievement of a heterogeneous group to a program that serves a specific language learner.

A second limitation to the research is that it relies on assessments in reading that begin collecting achievement data at the third grade. This is par for the course since state

and national assessments are given to students starting at that grade. Assessments prior to third grade are less normed making it difficult to conduct comparative studies. However, it also excludes the impact of instruction that was done prior to the third grade. For language acquisition; those primary grades are crucial for language development and reading readiness. At kindergarten, it is difficult to separate a language deficiency from a reading deficiency, which may result in misidentifying an ELL student for the appropriate support (Klinger, Artiles, & Mendez Barletta, 2006). For struggling readers, not developing reading proficiencies early on in school can be subjected to a Matthews Effect; where support and intervention are progressively reduced (Stanovich, 1986). Therefore, evaluating achievement data starting at the third grade ignores underlying factors that could explain continuous poor reading achievement.

The third limitation is that the relationship between language acquisition and special education are excluded from the research on effectiveness. There is precedence regarding the relation assessment and identification for language learners and learning deficits (Cummins, 1984). There is also research that ELL students are disproportionately represented by language acquisition program (Artiles et. al., 2005 & Klinger et. al., 2005). Although exploring the relation between language acquisition and special education is outside the scope of this study, it is important that the research address the issue.

This study explores an opportunity to compare students from both programs that have similar social and economic experiences. Comparing two groups of students that essentially demonstrate the same reading ability, the 4-hour ELD block can be evaluated in isolation. The study used achievement scores gathered at the primary level in attempts

to provide results that can be used to inform practices at earlier stages of development.

Conclusion of the Chapter

There is limited research regarding the impact of the 4-hour ELD block. State and national level data show that the reading achievement gap for ELL students is not decreasing, supporting a hypothesis that the 4-hour ELD block has failed to make a significant impact. In terms of this study, Hispanic ELL students are not succeeding at the same rates as their non-ELL peers. Not considered in the interpretation is that state and national reading results for the Hispanic ELL group include students that are enrolled in both English acquisition programs and the regular reading program, common among schools with large minority populations. Also not considered is how poverty affects reading achievement significant because Hispanic ELL students tend to be from poor communities. The study looks to control for these limitations in order to evaluate the impact of the 4-hour ELD block in isolation.

Chapter 3: Research Design

Introduction to the Study

The focus of this study is to measure the impact of the 4-hour ELD block by comparing two groups of students that come from the same social and economic community. A quantitative study with an experimental design was used as the most appropriate research method for this study. The study is quantitative because only numerical data is being examined. It is experimental because the study compares two similar groups with the first group receiving a treatment and the second group being a control. The data used consisted of R-CBM assessment scores at the time that all students started their prospective programs through the five years of the study. These reading assessment scores were evaluated for each year using various measurement tools as explained in this chapter. The results of each year were collected and compared in a longitudinal study to determine the consistency of results over time. AIMS reading results were also recorded to evaluate how participants in the study achieved on the state exam. The AIMS reading results were used to compare findings with current research using AIMS.

Research Design

The study used quantitative data, which consisted of existing reading assessment scores using an “in-district” benchmarking tool for reading. The scores were aggregated for the time period between Fall 2008 through Spring 2012, a five-year span. There were two independent variables for the study consisting of participants from two different reading programs. The first group consisted of limited English proficient (LEP) students that were enrolled in the 4-hour ELD block. The second group consisted of full English

proficient (FEP) students that were enrolled in the regular curriculum. The dependent variable in the study was the reading-curriculum based measure (R-CBM) scores collected for each participant. Data collection was done using a “between-subjects” study design. Qualified staff from the school district administered the R-CBM and reported scores electronically using the AIMS-Web® online product during the years of the study. The scores were attained by requesting information from district personnel via an excel spreadsheet.

The evaluations of reading scores were chosen with the assumption that, at the elementary level, learning reading and language acquisition are similar. The curriculum of the 4-hour ELD block is based on learning vocabulary, phonics, grammar, and reading comprehension, skills similar to the curriculum in the regular reading program. The research assumes that ELL students and non-ELL students are developing the same skills necessary for learning how to read as well as acquiring a first language. Therefore it is deduced that the R-CBM scores for students in the 4-hour ELD block and those in the regular reading program reflect a direct effect of the program the group was enrolled in.

It is recognized that reading scores can also be influenced by factors such as quality of instruction, leadership, teacher experience, teacher training, student’s attitude toward learning reading, etc. These factors could not be eliminated but were limited through the selection process. Participants in a unit of study are from the same school to ensure that the students received similar academic experiences that affect student achievement.

Students that attended the same school were chosen for each cohort in order to minimize factors based on teaching staff. Multiple schools from the district were

represented to get a cross section of varying school cultures. The research tests were conducted for each year of the study creating a longitudinal analysis of results over a period of time. The results over time helped to compare results consistent with continued progress over time.

Participants

The students participating in the study were identified using specific criteria during the selection process. Each criterion was applied in rounds using a process of elimination until a single representative for each cohort remained for each group in the study. As the cohorts were identified, there were incidents where there were more than two students in separate programs being represented. A randomization process was used to make sure two students were identified objectively. All other information collected during the selection process was used to verify the criteria but was not used as part of the statistical analysis. Any identifying information was coded after the selection process and prior to applying any statistical analysis.

All participants in the study were chosen from one school district where all schools in that district have similar social and economic communities. Studying only one school district will allow for the research to minimize, if not contain, variables that can contaminate a clear analysis on causality. All schools in the district have a high ELL population and are located in communities of poverty. It is assumed that participants had similar life experiences that would affect their access to learning and their progress in academic achievement.

All schools in the district had to adhere to the same state standards for reading instruction regardless of the program students were enrolled in during the study.

Although individual schools may have used different teaching materials and/or teaching practices, the state standards-based curriculum increased the probability that all students were taught the same information tested on the state assessment AIMS. The district employed a district level director that ensured teachers teaching the 4-hour ELD block were including the regular district curriculum for reading. The same district employed a director that ensured schools were implementing the district reading curriculum. The districts also required teachers in both programs to present lessons and were evaluated in instruction demonstrating the use of the regular curriculum. Having directors that reviewed and supported the different programs assumes that the quality of the materials, resources, and teacher training were equitable for all schools during the period of study.

Requirements for participation in the study. Students chosen for the study were based on specific criteria through process of elimination that served as a form of randomization. A criterion was applied in stages until a unit of study was identified for each school in each cohort. A randomization process was considered; however, a need to reduce dropout over the period of the study determined randomization would not be the best fit. The elimination process allowed the identification of students that would remain for the entire period of the study without subjecting a bias in the selection process. By nature of the process, the final data set would be reduced by each step.

Each unit of study is made up of two students each representing the LEP and FEP group. Each participant in the unit started the study with the same WPM score for the Spring R-CBM measure beginning in the first grade. The measure for spring is the first measure of the year and is conducted within two weeks after school starts. Using this score assumes that both students in the unit were beginning the study having the same

reading ability.

The study began with first grade R-CBM scores instead of kindergarten because this is the first year of compulsory education. Although it may be assumed that students received a structured reading program in kindergarten, it cannot be assumed that schools focused on a reading or language program that focused on an adopted curriculum. Therefore, starting with first grade scores assumes that students representing each unit had similar learning opportunities that resulted in the same starting reading score..

During the selections process it was discovered that for any given reading score, there were significantly more students representing the LEP group. Trying to reduce these groups into two representatives proved concerning; using an average or median score would render the data unreliable. Therefore, in cases where there were more representatives in one group, students were identified using a randomization process. Students' last names were organized alphabetically at the school level and participants were chosen in cardinal order. All remaining names were eliminated. Using this process, there were schools that had more than one unit of study.

Quantitative Data for Analysis

I chose reading fluency scores recorded for the Reading - Curriculum Based Measure (R-CBM) on the AIMS-Web® reading assessment. Reading fluency scores are used to determine a student's reading progress based on a normed scale for each grade level. Each student's reading score is measured by fluency or the correct number of words read per minute on a normed grade level passage. The student reads three passages and the student is given the best score of the reads. Analysts often refer to this score as the WPM score. The WPM score is measured three times a year, fall, winter, and spring.

Each assessment period has a norm-referenced score or “cut score” for each grade level. Cut scores are determined for each grade level using an algorithm that norms the reading level by factoring in the number of students taking the R-CBM for that assessment period. A student with a WPM score at or above the “benchmark” score is considered to be at a grade level for that particular time of the school year. Local analysts use the WPM score to determine instructional programming. By assessing students three times a year, local analysts can monitor a student’s reading progress and make changes to how instruction is provided.

Local analysts also use the WPM score to monitor program effectiveness by analyzing the percentage of students for each cut score at the grade and teacher level. Students meeting benchmark are said to be responding to their reading program effectively. For the instructional programming perspective, that means that if most students are making benchmark, the current reading program is considered effective and no changes or interventions are required. To support instructional programming, the assessment periods structure progress into three tiers, Benchmark, Strategic, and Intensive. When a majority of students are reading at the strategic level, it is common that programming changes are made such as increased teacher training, professional development, or class-level interventions. A reading specialist will conduct a more focused intervention program for intensive students. Analysts in charge of programming view a positive correlation between reading at benchmark and passing the AIMS test. It is common practice to use benchmark scores towards predicting the percentage of students who will pass the AIMS test.

The assessment also measures accuracy by dividing the number of words read correctly by the number of words attempted. The accuracy score is measured by a percentage. A score nearing 95% or better accuracy increases the reliability of the WPM score. This study will not utilize the 95% accuracy score because the process in identifying participants for the study and recording scores for the duration of the study makes incorporating the 95% accuracy score too complex. This is to say that a participant's accuracy score may vacillate greatly during the study and it would not be practical to invalidate or remove a participant's data considering the limited number of participants. Although the accuracy score would support the reliability of the WPM scores used in the study, it will not necessarily impact the finding of the study (e.g., students with low WPM scores may also have low accuracy scores.). Analyzing the accuracy scores could be the focus of a second study to validate the findings.

There are other similar tests such as Dibels® and the Northwest Evaluation Association or NWEA® that also measure reading ability using benchmark scores. The AIMS-Web assessment was chosen based on the researcher's knowledge and experience with this particular assessment tool. It is recommended that researchers with knowledge in other reading benchmark assessments should conduct the same study using the tool they are best familiar to compare results.

Process for Identification of Participants

Participants for the study were determined through a process of elimination using existing data. The district provided access to a spreadsheet of AIMS-Web data for every student enrolled in the district for the five years of the study. The data provided student identification by the following: student names, student identification numbers, gender

identification, birthdate, teacher of record for each school year, English language learner status, and Title I status. The spreadsheet also provided all measures used in the AIMS-Web assessment for each student. These measures included assessments on phonemic awareness, letter fluency, word fluency, and decoding. The use of these assessments was considered, but students are not tested in these areas after the second grade. Fluency is the only assessment that is assessed first through fifth grade.

The data was processed to meet three specific requirements to create the unit of study that would allow for the best analysis of student data. The first requirement was to have a unit of study that was made up of a student representative from the ELP and FEP groups. The unit of study would set the basis for the control variable in the study. The second requirement was to create a unit of study with both representatives from the same school. This focus helped to eliminate discrepancies of instructional quality between school dichotomies. The third requirement was to identify students who were consistently enrolled in the same school for the duration of the study. The intent was to eliminate dropout, a common challenge in longitudinal studies. The best way to achieve the intended unit of study was to apply a process of elimination when processing the available data. Student names, gender, and student identification numbers were used to validate criteria and student scores. Student identification information was also used to attain achievement scores for AIMS and AzELLA as well as to identify and eliminate participants enrolled in Special Education.

The first two levels of aggregation focused on meeting the requirements of the study. The first level was to look for students who were enrolled in first grade at a district school beginning in school year 2008. These students were aggregated from the

original spreadsheet and transferred to a new one. The original spreadsheet was saved and placed on a thumb drive for safekeeping. The second level was to then identify all students who maintained enrollment in the same district school for the 2008 school year through 2012. This information was then aggregated to a new spreadsheet with the old spreadsheet saved to the thumb drive.

The next level of aggregation focused on paring the data to units of study for each district school. The students remaining were grouped by their Spring 2008, RCB-M score for each district school by running a sort to line up scores in ascending order and running a filter to group by school. Any student that did not have a matching score with another student in their school was then removed from the spreadsheet. The student groups were reviewed to determine if they had at least one representative from the LEP and FEP classification. Those student groups that did not have both representatives were removed from the spreadsheet.

The final level of aggregation looked to reduce student groups to create units of study. As it will be discussed in the findings, the student groups were skewed with many more students representing the LEP group. Supplemental information was requested for the remaining students to address the imbalance of representation and reduce the numbers of students to two. Student data identifying special education was then used to eliminate any remaining students that were enrolled in special education. The reasoning behind that was that students with learning disabilities could have not done well in a reading program because they were not mentally capable of progressing at an average rate and not because the program made a negative impact.

AIMS reading achievement scores for the remaining students were also requested. The data was not used to identify participants but will be used to support analysis in the findings and final recommendations. The aggregation of student group data was considered as a way to measure effect size; however, the imbalance in representation would not be statistically efficient for a t-test and therefore the consideration was abandoned.

Statistical Analysis of Existing Data

Three statistical analyses were used to report findings regarding the research questions. A student T-Test was used to identify any statistical difference between the two programs. Effect Size was used to determine rate of growth for individual participants in order to identify differences between the two programs. Finally, reading achievement scores for AIMS were used to compare the percentage of students who pass the AIMS from each program for grades three through five. The student T-Test is the main analysis to identify the impact of the 4-hour ELD block. The Effect Size analysis was used to measure magnitude of impact. AIMS reading scores were used to connect this study to prior research using AIMS test results.

Student t-test. An independent t-test was used to compare the Spring WPM scores for participants in the LEP group against the participants in the FEP group. The WPM score is based on the average number of words the participant can read in one minute given three normed, grade-level reading passages. The passage is provided by the AIMS-Web® program. The Spring WPM score represents the final outcome achieved by participants after a year of reading instruction. By comparing the difference in means of

Spring WPM scores between the LEP and FEP groups, it can be determined if the LEP group achieved WPM scores that were significantly different than the FEP group after receiving instruction in their respective reading programs.

The independent t-test applied a null hypothesis assuming that no significant difference existed between the Spring WPM scores for the LEP and FEP group. The alternate hypothesis assumed that there was a significant difference between the Spring WPM scores for the LEP and FEP group. The null hypothesis was accepted if the results of the independent t-test yielded a p-value more than 5%. Accepting the null hypothesis assumed that the participants in the 4-hour ELD block achieved significantly higher or lower WPM scores on average than the regular reading program. Accepting the alternate hypothesis assumed that participants in either the 4-hour ELD block or the regular reading program achieved similar WPM scores. Accepting the alternate hypothesis can also be interpreted as participating in either reading program would result in the same outcome.

If the results of the independent t-test demonstrated there was a significant difference, it was assumed that the 4-hour ELD made an academic impact for the participants in the LEP group. However, the independent t-test did not provide information identifying the magnitude of any significant difference. Results did not reveal if the LEP group received an advantage or disadvantage by being in the 4-hour ELD block. An effect size measurement was used for further testing to address the limitation.

Applying the independent t-test to one year of data provided information on that year alone. Many variables such as teacher ability, leadership, and training can affect

learner outcomes in any given year. Therefore, the t-test was conducted for each year of the study in order to analyze data overtime. The data over time measured for any possibility that participants in the LEP group yielded different results the longer they received instruction in the 4-hour ELD block program.

Effect size. An effect size was calculated to measure the difference between the Fall WPM score and Spring WPM score for each program group. The Fall WPM score is based on a similar leveled reading passage as used for the Spring WPM score to ensure students are not memorizing a reading passage. The passages are provided by the AIMS-Web® program. The Fall WPM score represents the reading ability at the start of a given school year. The Spring WPM score represents the reading ability achieved at that given year after the reading program or treatment was given. In this respect, the Fall WPM score is considered a “pre-test” while the Spring WPM is considered a “post-test”. It is assumed that the 4-hour ELD block made a positive impact if the participants acquired at least one year’s growth or an effect size of .40 or higher. Conversely, it is assumed that the regular reading program made a positive impact if the participants acquired at least one year’s growth or an effect size of .40 or higher. Therefore, the effect scores for the FEP group were recorded in order to make a comparative analysis with the LEP group and enhance findings for measuring the impact of the 4-hour ELD block.

The effect size was calculated by applying a paired t-test and the Cohen’s d formula. The results were interpreted using Hattie’s scale for educational effect size (Hattie, 2009). A paired t-test was conducted to measure the difference between the Fall WPM scores and Spring WPM scores for a given program group. The paired t-test provided the mean (M) and the standard deviation (SD) for the two sets of scores. The

Cohen’s d formula was then applied to calculate effect size by dividing the standard deviation from the mean.

Cohen’s d:

$$effect\ size = \frac{Mean\ of\ the\ program\ group}{Standard\ Deviation\ of\ the\ program\ group}$$

The effect size score was then translated into an equivalent measure of time in respect to expected growth using John Hattie’s educational scale for effect size as shown in table 1.

Table 1

<i>Effect Size Translated into Measured Annual Growth</i>	
Effect Size	Growth over a 12-month period
.15	No growth
.2	9 months growth
.4	12 months growth
1.0	2-3 years growth

The effect score was paired with a p-value provided by the paired t-test, which identifies if the difference between the pretest and post-test was significant. A significant difference between scores assumes that the students in the program responded to the instruction in reading program. No significant difference would indicate that the instruction during the reading program had no effect towards changing achievement scores. The effect size was then measured to determine if the change was positive or negative and how much of a change was made.

The effect size for the LEP group was also compared to the effect size of the FEP group. The difference in magnitude between the LEP group and FEP group was used to determine whether the LEP group experienced an advantage or disadvantage by being in the 4-hour ELD block. A greater effect size for the LEP group assumed the LEP participants experienced an advantage by being in the 4-hour ELD block. Adversely, a

lesser effect size for the LEP group assumed the LEP participants experienced a disadvantage by being in the 4-hour ELD block. If no significant difference was determined for either program group, the comparison was assumed to be invalid and any analysis was inconclusive.

As with the independent t-test, calculations for effect size were conducted for each year of the study. The data recorded over time was measured to see if participants in the LEP group yielded different results the longer they received instruction from the 4-hour ELD block. The comparison of effect size for the LEP and FEP groups were graphed to analyze any trends that would support any conclusions that were made.

AIMS results. The results of the AIMS reading test for the LEP group were compared to the results of the FEP group. The AIMS results are based on a passing score of “Meets” or “Exceeds” on the Language Arts portion of the AIMS test. The AIMS test is only given to students in third through fifth at the elementary school level; which is represented by the third through fifth year of the study. Therefore, AIMS results are only recorded for those three years of the study. A passing score on the AIMS reading test assumes that the participant has acquired the expected grade-level content and is able to apply that content effectively after a year of instruction. A passing AIMS score also assumes that the participant is functioning academically at grade-level.

The percentage passing AIMS was calculated by dividing the number of participants in a program group with an “Exceeds” or “Meets” by the number of participants in that group. The percentage passing AIMS for both groups was recorded on one graph along with the percentage of members failing in each group. The graph represents all participants in the study with all percentage parts equaling a total of 100%.

A percentage of participants passing the AIMS reading test was measured to analyze the results of programing in regards to state testing. The percentage was calculated by dividing the total number of participants in a program group from the number of participants of that group passing the AIMS reading test. The percentages were then used to compare program groups at the ability level as well as a whole. The measurement was applied for the last three years of the study; grades three through five were the only grades that produced AIMS scores. The analysis for this measure was used to support the finding from the t-test. The AIMS reading analysis will also attempted to relate my findings with current research using AIMS reading results to analyze the effectiveness of the 4-hour ELD block referenced in Chapter two. The AIMS findings may also be included as part of a presentation for any requests made by the school board of the district were the study was conducted.

Data Gathering Process

Cartwright School District was chosen for the study because it met the demographics specific to the study and results would benefit the district directly. Cartwright School District is located on the west side of the Phoenix metropolitan area in Arizona known as Maryvale. According to the 2015 statistics on Datausa.io, Maryvale has a median household income of \$36,618, which is comparably lower than the \$50,000 reported for the whole state. Maryvale reports that 32.9% of the population has poverty status, which is higher than the national average of 14.7%. Hispanics or Latinos make up approximately 49% of the Maryvale population in poverty. The census data indicates that Maryvale meets the requirement for being a community of poverty. According to the same website, the most common ethnicity in Maryvale is Hispanic (approximately 76%

of the population) and 64% of the Maryvale population speaks a language other than English. The census data indicates that Maryvale is predominately Hispanic with Spanish the most commonly used language. Most LEP students attending Cartwright School District have Spanish as their primary home language. Datausa.io aggregates data from information reported to the Census Bureau.

Cartwright School District is made up of 20 schools ranging from kindergarten to eighth grade with three structured variations. Twelve of the schools offer grades Kindergarten to fifth, four schools offer grades kindergarten to eighth, and four schools offer grades sixth to eighth. Cartwright School District had 19,191 students during the start of the study. The student population decreased to 16,541 by the end of study. Approximately 95% of the student population was Hispanic with 59% enrolled in the 4-hour ELD block. Cartwright School District had approximately 93% of their students receiving free or reduced lunch.

Research timeline. Permission was requested to review archived data from the Cartwright School Board at a regularly scheduled board meeting in the fall of 2017. Permission was granted that night after a presentation regarding the purpose of the study was given. The researcher was directed by the superintendent to work with Cartwright's Assistant Director for Assessment and Data Management to retrieve archival records. The Assistant Director provided a spreadsheet based on a query request that was provided to his office. The official request was presented the month of February 2018. The data gathering process was concluded by June of 2018.

The first round of data aggregation was to attain a list of all students who enrolled in first grade at a Cartwright school beginning in the fall of 2008 and who reported a Fall

R-CBM score. The Assistant Director and assistant provided AIMS-Web® data for the school year 2008. They also provided data from Synergy, a state enrollment data platform, which included a list of students that were enrolled at the start of the same school year. The data also indicated whether or not a student was enrolled in the SEI program that year. A spreadsheet was created to aggregate data of the students who were in first grade and had a valid Fall R-CBM score. Students were first organized by school then by WPM score. Students with common scores were identified as a cohort. Any school cohort that did not have more than one representative from the LEP or FEP program group was disqualified.

The second round of data aggregation was to identify students that attended the same school in the Cartwright School District for the period of the five-year study. In April of 2018, the spreadsheet of aggregated data was provided to the Assistant Director requesting a search based on continuous enrollment for the students identified in the first round. The Assistant Director provided a Synergy report of all students that were enrolled in the same school for the five years of the study. The Synergy list and the spreadsheet were correlated to determine which students on the spreadsheet maintained continuous enrollment at the same school for the full period of the study. Students who were not continuously enrolled were disqualified. The data for disqualified students were removed from the spreadsheet. The remaining school cohorts were then reviewed to ensure there was representation from each program group. School cohorts that did not have the minimum representation were also disqualified and their data removed from the spreadsheet.

The third round of data aggregation was to record R-CBM and national percentile scores for each student. In May of 2018, an updated spreadsheet was presented requesting the Fall and Spring R-CBM and national percentile scores for each year of the study. The Assistant Director provided a AIMS-Web data report for all measures from their AIMS-Web database. That data was then aggregated from the report and added to the spreadsheet. Based on the data given, not all students on the updated spreadsheet had valid scores for each interval recorded. This was not a frequent occurrence and it was decided to keep students with missing scores in the study mainly to maintain a viable number of participants. It was recognized that missing scores could affect the calculation for effect size when testing for that given year data was missing. In these occurrences, the scores were left as “not applicable” and recorded into the statistical calculation.

Adjustments to the process. An issue emerged regarding learning impairments. In some cases, participants that qualified in all rounds had scores that did not increase appropriately from year to year. The scores for these participants were analyzed further and it was discovered that in most cases, the participants never progressed beyond the 10% national percentile ranking. It is common to use the 10% percentile rank as an indicator for possible learning impairments. Therefore, the data was removed from the data set because it was assumed that the participants were receiving different reading instruction and the R-CBM scores were more indicative of a learning impairment. On further review, the occurrences were mostly applicable to students in the FEP group. To ensure an unbiased selection process, additional information regarding special education services was requested from the district. All participants that were identified for services in Special Education for reading were removed from the data set.

The original proposal for the study was to measure the effect size for each participant and use a graph to analyze the difference between the two program groups. It was my intention to measure annual growth using R-CBM scores and national percentile rankings. As the data was graphed, no apparent trends emerged to suggest a conclusive measure of impact. The data did seem useful in measuring how well each individual student responded in their respective program. The national percentile ranks were then abandoned and the use of the t-test was determined to be a better tool to statistically measure impact.

Using a t-test exposed a problem with the data set collected. Almost all school cohorts did not have equal representation for each program. Although this would not affect measurements for effect size, it did create analysis complications for the t-test. In many cases, there were many more students in the LEP program group than in the FEP group; often the FEP group was represented by only one participant. To adjust for the imbalance, the school cohorts were reduced to a unit of study with one participant from each program group being represented. To avoid having to restart the selection process or rely on a randomizer tool, the spreadsheet was reorganized and a uniform selection method was eliminated to bias.

During the data gathering process, it was brought to my attention that the study was lacking AIMS scores as part of the analysis. AIMS scores would be relevant to school practitioners reviewing the study and it would be prudent to present an AIMS analysis to the school board upon its completion. AIMS data would also connect this study to existing research. Therefore, an additional request for AIMS data was made to the Assistant Director in spring of 2018 and was included in the study.

Bias and Ethical Considerations

One bias to consider is that the researcher was a principal in the district being studied. The school the researcher was an administrator at is represented more than once in the data set. To ensure a non-biased process, the researcher chose to use existing data that could not be manipulated because it was recorded at the time the assessment was given. Names of the students and schools were also coded during calculations to limit knowledge of which schools produced specific results. The researcher no longer works for the district, so any bias towards influencing district practice and policy is also minimal.

A second bias to consider is the strong opinions of the researcher regarding meritocracy and the effects of poverty on academic achievement. Knowledge from academic programs and different professional development activities has shaped the researcher's approach to teaching students through the lens of poverty. The research also brings over twenty years of knowledge and experience in the teaching field working with the demographics addressed in the study and over thirty years of formal study and training. It is difficult to separate what was learned during those years from empirical and anecdotal research. For this reason, only quantitative data was used in order to preserve unbiased results. The goal was not only to complete a dissertation but also to learn from the study and improve practice. That being said, it should also be acknowledged that the researcher is not proficient in statistics. The analysis of the data is based on how the researcher has evaluated educational data for making instructional decisions in the past while in the capacity of being a teacher and administrator. This is an applied research project and the main purpose of this paper is to inform practitioners.

Limitations and Potential Benefits of the Study

This is a very focused study using a relatively small sample. The interpretation of the results is limited to urban school demographics with high populations of poor and ELL students. The study does not address rural areas that may also have the same demographics but not as many students or schools in affluent communities. It also does not address the impact of the 4-hour ELD program on ELL students who are proficient in their first language. These students are more likely to enter the 4-hour ELD program at later grades. Further research will be needed to address these different demographics. Furthermore, the tool used in the study, AIMS-Web®, is not a standard tool considered to be empirical research. An analysis can only be made based on what the program was designed to do. It will be up to other researchers to accept the assessment data it produces as viable. Therefore, the study cannot provide a complete analysis on the impact of the 4-hour ELD block for all schools that are implementing it.

Only seven of the twenty schools in the district are represented which may not provide substantial evidence for the district. Considerations were made when creating cohorts by reading ability only, effectively making larger cohorts with participants from different schools. However, principals and teachers have some autonomy on how materials and resources are utilized on their campus, which students receive additional academic support, and determine which teachers will teach the different programs, making some variables difficult to control. Using the five-year period allows for more consistent results.

Teacher efficacy is also a variable that could affect results. Teachers have

different levels of knowledge, experience, and training that effect how the curriculum is presented and how students respond to the information. By looking at five years of data, all participants in the study, from either sub-group, have the same opportunities to have instructional experiences that positively or negatively affect achievement. Although this is less than half the schools, I felt it prudent to maintain the sample size even if not all schools are represented and some schools were represented more than once.

It is also recognized that selecting students that were continuously enrolled in the same school can also affect achievement results. A hypothesis commonly accepted in the field is that students who are continuously enrolled tend to achieve well on state testing. However, it is not common for schools to have a high percentage of students that are continuously enrolled. This requirement resulted in examining scores for students with continuous enrollment, which is not indicative of the school population.

The reading benchmark tool used by the district is a “for-profit” product sold to the district. The validity and interpretation of the scores are solely dependent on the formulas and scales used by the company. Although there is no indication that the tool is biased towards any specific programming, it is prudent to disclose that the product is designed and sold by the same company that creates the state testing. It was an existing tool staff and students had experience and familiarity with. The tool was chosen to eliminate biases inherent when participants know they are being subject to research and to eliminate any intentional or unintentional recording from administering the tool incorrectly.

It is important to note that students in the LEP groups were not continuously

enrolled in the 4-hour ELD block for the duration of the study. That means that some of the LEP students were receiving instruction using the regular reading program in the later years of the study. Since the longitudinal study was designed to follow a participant's progress, LEP students remained in the group even after they exited. The assumption is that although a student exited the 4-hour ELD block, it was due to the student's success in the program. Information regarding student exit dates were collected but not used. Further study on the effects of student progress after exiting the 4-hour ELD block would be excellent as a further research project.

Three reading ability groups emerged during the process of identifying participants but were not used for analysis. The first group was made up of participants who entered the first grade reading zero words per minute. The zero score indicated these participants had little to no reading ability commonly known as pre-readers. The second group had participants with scores between one and five words per minute. Scores in this range are approaching the fall R-CBM of six. Students with scores in this range are considered emergent readers. The third group had participants with scores of six or more words per minute. Students with scores in this range are considered reading at grade level. The current data uses the average score for the cohorts combined. Findings of the data do not define how the 4-hour ELD block impacted different reading abilities. The cohorts were not used because the selection process reduced cohorts to small sample sizes that invalidate any statistical analysis. It would be beneficial for a study to be conducted using larger sample sizes for each cohort.

The potential benefit the study does provide is the ability to isolate the program in

order to measure its impact. This study may provide researchers a glimpse on how the 4-hour ELD affects reading achievement through a controlled two-variable environment. Although the study is limited in scope, results may better define different conceptual frameworks when studying the 4-hour ELD block, education programming for poor communities, and measuring the impact of programs that have limited research. The results may then provide better support for the use of current hypotheses on academic programming in regards to language acquisition and intervention.

Conclusion of the Chapter

Current research on the impact of the 4-hour ELD block demonstrates that academic achievement in reading for ELL students have improved since its implementation but has not closed the achievement gap. The research does not detail how ELL achievement scores have improved or why the achievement gap continues to occur despite the use of the program. The study in this research was designed as a basic comparative study between a treatment group and a control group. The study used reading achievement scores from an existing benchmark tool, which recorded reading progress for both groups. The interpretation of the data assumes that the results are correlated to the impact of the program that was tested. The results should only be interpreted for demographics with high ELL populations and poor communities.

Chapter 4: Results and Analysis

Data Set

The data set has seven out of fifteen schools in the district, making up 47% percent of the district schools represented as shown in table 2. Twenty units of study were identified through the selection process, which makes forty participants completing the full data set. There are an equal number of participants representing the LEP and FEP groups with twenty in each group. The data for the student t-test will have an N-Count of forty providing a statistically relevant sample.

Table 2

Makeup of WPM scores for LEP and FEP Groups

Starting Score LEP Group	Starting Score FEP Group	School Attended
0 WPM	0 WPM	School 2
0 WPM	0 WPM	School 2
0 WPM	0 WPM	School 4
2 WPM	2 WPM	School 2
3 WPM	3 WPM	School 2
5 WPM	5 WPM	School 3
7 WPM	7 WPM	School 4
9 WPM	9 WPM	School 2
9 WPM	9 WPM	School 5
11 WPM	11 WPM	School 4
11 WPM	11 WPM	School 5
11 WPM	11 WPM	School 6
12 WPM	12 WPM	School 5
13 WPM	13 WPM	School 5
13 WPM	13 WPM	School 5
13 WPM	13 WPM	School 7
14 WPM	14 WPM	School 6
18 WPM	18 WPM	School 5
22 WPM	22 WPM	School 1
26 WPM	26 WPM	School 6

Three ability groups were identified during the selection process. The first group was the pre-reader group. There were six participants that made up the pre-reader group. The participants in this group scored zero words per minute at the start of the study. The second group was the emergent reader group. There were eight participants in this group.

The participants in this group scored at least one word per minute but less than seven words per minute at the start of the study. The third group was the at-grade level group. The third group was the most represented for the study with twenty-six participants. The participants in this group scored more than seven words per minutes at the start of the study. The ability groups will be used to organize effect size analysis and AIMS reading percentages.

Student T-Test

An independent student t-test was conducted to compare Spring WPM scores for participants in the LEP group and participants in the FEP group for each year of the study. The results were reported by each year of the study and will provided the n-score used to calculate the score as well as an analysis of the results. The null hypothesis is that there is no significant difference between the scores of the LEP group from the FEP group. The null hypothesis will be accepted if the test calculates a p-value more than .05. The alternative hypothesis is that there is a difference between the LEP and FEP scores for WPM. The null hypothesis will be rejected if the test calculates a p-value less than .05. A list of statistical abbreviations is provided in table 3. A summary of the analysis will follow

Table 3

List of Statistical Abbreviations

N-score	The number of participants included in the test.
T-value	The results of a t-test analysis that indicates where the value would be on a sampling distribution table. The t-value is used to calculate the p-value.
P-Value	The score that measures whether there is a significant difference between the groups tested. The p-value is used to test against the null hypothesis.
M	The mean for the set of scores tested.

SD The standard deviation for the set of scores tested.

Student t-test analysis for year one. Table 4 provides the t-test results that were used to find the p-value for our analysis. The results indicate that in year one of the study, there was not a significant difference between the scores for the LEP group (M=43.61, SD=24.61) and the FEP group (M=47.72, SD=25.48) conditions; $t(34)=-4.92$, $p=.626$.

Table 4

Independent T-Test Based on Spring WPM Scores for 2008 (Year 1)

Participation in Program	N	Mean	Std. Deviation	Std. Error Mean
In Program	18	43.61	24.605	5.800
Not in program	18	47.72	25.481	6.006

	F	Sig	t	df	Sig. 2-Tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances Assumed	.007	.936	-.492	34	.626	-.4111	8.349	-21.078	12.856
Equal Variances not Assumed			-.492	33.959	.626	-4.111	8.349	-21.079	12.857

A p-value of .62 is larger than .05 and therefore we accept the null hypothesis. This indicates that the students in the LEP group did not do any better or worse than the students in the FEP group. This could indicate that the 4-hour ELD block did not have a significant impact over the regular curriculum. The standard error difference of 8 points indicates a relatively small difference between the two groups.

Student t-test analysis for year two. Table 5 provides the t-test results that were used to find the p-value for our analysis. The results indicate that in year two of the study,

there was also not a significant difference in the scores between the LEP group (M=81.68, SD=25.42) and the FEP group (M=83.06, SD=31.71) conditions; $t(33)=-.143$, $p=.887$.

Table 5

Independent T-Test Based on Spring WPM Scores for 2009 (Year 2)

Participation in Program	N	Mean	Std. Deviation	Std. Error Mean
In Program	19	81.68	25.417	5.831
Not in program	16	83.06	31.708	7.927

	F	Sig	t	df	Sig. 2-Tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances Assumed	.453	.506	-.143	33	.887	-1.378	9.653	-21.018	18.261
Equal Variances not Assumed			-.140	28.637	.890	-1.378	9.841	-21.516	18.759

A p-value of .887 is greater than .05 and therefore we accept the null hypothesis. This indicates that students in the LEP program for two years still did not make a larger impact than those students who remained in the regular program. A standard difference error of 9.65 indicates a relatively small difference between the two groups but the difference is larger than in year one.

Student t-test analysis for year three. Table 6 provides the t-test results that were used to find the p-value for our analysis. The results indicate that in year three, there again appears to be no significant difference in the scores between the LEP group (M=104.70, SD=37.90) and the FEP group (M=101.94, SD=35.45) conditions: $t(36)=.231$, $p=.82$.

The p-value of .82 is larger than .05 and therefore the null hypothesis is accepted as well. This indicates that after three years in the program, the students in the LEP program still did not measure a significant difference in achievement. With a standard error score of 11.94, the difference is greater than year two indicating the range steadily increased over the period of the study.

It is important to note that this would be grade three in the study. This is the grade that begins state testing on the AIMS and could indicate a focus on reading instruction to achieve better success rates. It is also important to note that some of the LEP students are beginning to exit the 4-hour ELD program and are participating in the regular reading curriculum.

Table 6

Independent T-Test Based on Spring WPM Scores for 2010 (Year 3)

Participation in Program	N	Mean	Std. Deviation	Std. Error Mean
In Program	20	104.70	37.902	8.475
Not in program	18	101.94	35.448	8.355

	F	Sig.	t	df	Sig. 2-Tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances Assumed	.202	.656	.231	36	.819	2.756	11.944	-21.468	26.979
Equal Variances not Assumed			.232	35.939	.818	2.756	11.901	-21.382	26.893

Student t-test analysis for year four. Table 7 provides the t-test results that were used to find the p-value for our analysis. The results indicate that in year four, again the results indicate that there is no significant difference between the LEP group (M=120.50, SD=29.74) and the FEP group (M=108.84, SD=34.69) conditions;

$t(37)=1.13, p=.27$.

The p-value of .27 is greater than .05 and therefore we again accept the null hypothesis. The standard error of difference of 10.33 is not an increase over year 3 indicating the difference between the two years is not changing. However, the p-value is reducing in value. Although this indicates that there is not a significant difference in how students responded to either program, it does indicate that the students in the LEP group are possibly showing some benefit to being in the 4-hour ELD block.

Table 7

Independent T-Test Based on Spring WPM Scores for 2011 (Year 4)

Participation in Program	N	Mean	Std. Deviation	Std. Error Mean
In Program	20	120.50	29.735	6.649
Not in program	19	108.84	34.693	7.959

	F	Sig.	t	df	Sig. 2-Tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances Assumed	.218	.643	1.129	37	.266	11.658	10.329	-9.271	32.587
Equal Variances not Assumed			1.124	35.507	.269	11.658	10.371	-9.385	32.701

Student t-test analysis for year five. Table 8 provides the t-test results that were used to find the p-value for our analysis. The results indicate that in the final year of the study, there is no significant difference between the LEP group (M=142, SD=36.06) and the FEP group (M=133.90, SD=41.42) conditions; $t(38)=.660, p=.51$.

The p-value of .51 is greater than .05 therefore we accept the null hypothesis. The standard error difference of 12.28 does show an increase from the previous years in the study. Again, this indicates that the students in the LEP group did

not benefit more from the 4-hour ELD block than their FEP counterparts did in the regular reading curriculum.

The p-value also increases significantly with the larger changes happening in the later years of the study. This may indicate that other factors may have influenced scores over program benefits such a focus on standards and student attitudes toward schooling. In the later grades, students are typically working on standards that focus on reading comprehension and less on improving reading skills. Students in the later elementary grades who are still struggling with reading skills are receiving help during scheduled interventions. As students are getting older, they are preparing to move to middle school and it is possible that students begin to prioritize social interactions over improving academically.

Table 8

Independent T-Test Based on Spring WPM Scores for 2012 (Year 5)

Participation in Program	N	Mean	Std. Deviation	Std. Error Mean
In Program	20	142.00	36.063	8.064
Not in program	20	133.90	41.422	9.262

	F	Sig	t	df	Sig. 2-Tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal Variances Assumed	.599	.444	.660	38	.514	8.100	12.281	-16.761	32.961
Equal Variances not Assumed			.660	37.293	.514	8.100	12.281	-16.776	32.976

Summary of Analysis

Based on the results of the t-test, it is prudent to accept that the 4-hour ELD block did not make a significant difference for LEP students when compared to their English

proficient peers. This indicates that the LEP students R-CBM scores were similar to the FEP group for each year of the study. Since the scores are similar, an interpretation can be made that the LEP students may have scored significantly poorer than FEP students if they were not in the 4-hour ELD block.

It is important to note for further reference that the greatest variation in scores were in the first year of the study. This follows the hypothesis that in the first grade, learning a language and learning to read are very similar. A difference in instructional programming begins to diverge further along in the grade level. Testing the hypothesis is beyond the scope of this study but would be a suggestion for further research.

Effect Size

The effect size was used to measure the rate of growth participants in an ability group demonstrated by increasing their WPM score from the Fall to the Spring R-CBM assessments. These rates were compared between the LEP and FEP groups for each ability group. Two tests were conducted in order to find the effect size for each ability group. The first test was a paired sample t-test using the SPSS program to find the mean and standard deviation. The mean and standard deviation for each test was used to apply Cohen's d formula to measure effect size by dividing the mean by the standard deviation. Hattie's scale for effect sizes (Table 1) was used to interpret the results of the test. The calculations to find effects sizes were conducted for each ability group and for each year of the study. A summary of analysis will follow.

Effect size analysis for year one. In year one of the study, a paired samples t-test was conducted for the LEP group with $M=30.56$ and $SD=20.55$ conditions; $N=18$,

p=.003. This results in an effect size of 1.49 for the LEP group. A p-value of .003 indicates there was a significant difference between the fall and spring R-CBM scores. This indicates that students benefited greatly from being in the 4-hour ELD block. An effect size of 1.49 would indicate that LEP students made on average a year's or more growth in the first grade.

The paired samples t-test conducted for the FEP group resulted in M=37.67 and SD=20.38 conditions; N=18, p=.00. This results in an effect size of 1.85 for the FEP group. A p-value of .00 also indicates that there was a significant difference between the fall and spring R-CBM scores. This would indicate that the students benefited greatly from being in the regular reading program. An effect size of 1.85 indicates that the FEP students made on average a year's or more growth in the first grade (see table 9).

Table 9

Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2007 (Year 1)

	Mean	Std. Deviation	Std. Error Mean	Paired Differences		t	df	Sig. 2-tailed
				99% Confidence Interval of the Difference				
				Lower	Upper			
LEP Group	33.556	20.552	4.844	23.335	43.776	6.927	17	.000
FEP Group	37.667	20.382	4.804	27.531	47.802	7.841	17	.000

The data for both the LEP and FEP group indicates that students made significant gains in the first grade. Students in both groups benefited from the program they enrolled in. The FEP group had a slightly higher effect score with a difference of .36 points higher. Translated in yearly growth, this could be interpreted to mean that the FEP group grew almost a year more than their LEP peers.

Effect size analysis for year two. In year two of the study, a paired samples t-test was conducted for the LEP group with M=37.83 and SD=19.91 conditions; N=18, p=.001. This results in an effect size of 1.90 for the LEP group. A p-value of .001 also indicates there was a significant difference between the Fall and Spring R-CBM. This indicates that students benefited significantly from being in the 4-hour ELD block. An effect size of 1.90 indicates that LEP students made on average a year's or more growth in second grade. The effect size is also larger in the second year than in the first year indicating a positive trend towards increasing reading skills.

The paired samples t-test conducted for the FEP group resulted in M=39.25 and SD=29.44 conditions; N=16, p=.00. This results in an effect size of 1.33 for the FEP group. A p-value of .00 indicates that there was a significant difference between the Fall and Spring R-CBM scores. This indicates that students benefited greatly from being in the regular reading program. An effect size of 1.33 indicates that FEP students on average made a year's or more growth in the second grade. For the FEP group, the effect score is lower than the first year of the study indicating a negative trend towards increasing reading scores (see table 10).

Table 10

Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2008 (Year 2)

	Mean	Std. Deviation	Std. Error Mean	Paired Differences		t	df	Sig. 2-tailed
				99% Confidence Interval of the Difference				
				Lower	Upper			
LEP Group	37.833	19.912	4.693	27.931	47.735	8.061	17	.000
FEP Group	39.250	29.436	7.359	23.565	54.935	5.334	15	.000

In the second year of the study, students in either the LEP and FEP programs are making significant gains while being in their respective programs. In this year, the LEP

group makes a larger average gain with a difference of .57 points in growth. Translated in yearly growth, the LEP group is making more than a year's growth over their FEP peers.

Effect size analysis for year three. In year three of the study, a paired samples t-test was conducted for the LEP group with $M=39.58$ and $SD=16.32$ conditions; $N=19$, $p=.00$. This results in an effect size of 2.43 for the LEP group. A p-value of .000 indicates there was a significant difference between the fall and spring R-CBM scores. This indicates that students benefited significantly from being in the 4-hour ELD block. An effect size of 2.43 indicates that LEP students made on average made over two year's growth in third grade. The effects size in year three continues to increase from previous years following the positive trend in improving reading scores.

The paired samples t-test conducted for the FEP group resulted in $M=40.11$ and $SD=13.03$ conditions; $N=18$, $p=.00$. This results in an effect size of 3.08 for the FEP group. A p-value of .000 indicates there was a significant difference between the Fall and Spring R-CBM. This indicates that students benefited significantly from being in the regular reading program. An effect size of 3.08 indicates that FEP students gained more than one year's growth in third grade. The effect score for year three increases greatly from the previous years of the study. It not only reverses a negative trend beginning in year two but also increases higher than year one. In effect, it seems to negate the dip observed in year two (see table 11).

Table 11

Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2009 (Year 3)

	Mean	Std. Deviation	Std. Error Mean	Paired Differences		t	df	Sig. 2-tailed
				99% Confidence Interval of the Difference				
				Lower	Upper			
LEP Group	39.579	16.317	3.743	31.714	47.444	10.573	18	.000
FEP Group	40.111	13.029	3.071	33.632	46.590	13.062	17	.000

In the third year of the study, students in either the LEP and FEP programs are making significant gains while being in their respective programs. In this year, the FEP group makes a larger average gain with a difference of .65 points in growth. Translated in yearly growth, the FEP group is making more than a year's growth over their LEP peers. It seems that students in the FEP group recoup growth in reading scores although indicators are not clear as to why. One thing to note is that third grade is the start of state testing, which may be a factor for how reading is being taught and the importance it is getting in attention to programming.

Effect size analysis for year four. In year four of the study, a paired samples t-test was conducted for the LEP group with $M=27.45$ and $SD=10.90$ conditions; $N=20$, $p=.00$. This results in an effect size of 2.52 for the LEP group. A p-value of .000 indicates there was a significant difference between the fall and spring R-CBM. This indicates that students benefited significantly from being in the 4-hour ELD block. An effect size of 2.52 indicates that LEP students continued to make more than one year's growth in fourth grade. The LEP group continues to demonstrate an upward trend from past years in the study. Further study will be needed to determine how consistent that growth has been.

The paired samples t-test conducted for the FEP group resulted in $M=18.84$ and $SD=14.97$ conditions; $N=19$, $p=.00$. This results in an effect size of 1.26 for the FEP group. A p-value of .000 indicates there was a significant difference between the Fall and Spring R-CBM. This indicates that students benefited significantly from being in the regular reading program. An effect size of 1.26 indicates that FEP students gained more than one year's growth in fourth grade. The effect size for the FEP group once again decreases significantly than year three (see table 12).

Table 12

Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2010 (Year 4)

	Mean	Std. Deviation	Std. Error Mean	Paired Differences		t	df	Sig. 2-tailed
				99% Confidence Interval of the Difference				
				Lower	Upper			
LEP Group	27.450	10.899	2.437	22.349	32.551	11.263	19	.000
FEP Group	18.842	14.968	3.434	11.628	26.056	5.487	18	.000

In the fourth year of the study, students in either group continue to demonstrate significant growth in reading scores. Students in the LEP group make a larger average gain in effect size by a difference of 1.26 or twice as much as the FEP group for the same year. The LEP group continues to make gains in effect size while the FEP group reverts to a decrease in average scores.

Effect size analysis for year five. In year five of the study, a paired samples t-test was conducted for the LEP group with $M=34.80$ and $SD=16.40$ conditions; $N=20$, $p=.00$. This results in an effect size of 2.12 for the LEP group. A p-value of .000 indicates there was a significant difference between the fall and spring R-CBM scores. This indicates that students benefited significantly from being in the 4-hour ELD block.

An effect size of 2.12 indicates that LEP students continue to gain more than one year’s growth in fifth grade. The effect score for year five begins to decrease slightly.

The paired samples t-test conducted for the FEP group resulted in $M=27.65$ and $SD=15.01$ conditions; $N=20$, $p=.00$. This results in an effect size of 1.84 for the FEP group. A p-value of .000 indicates there was a significant difference between the Fall and Spring R-CBM. This indicates that students benefited significantly from being in the regular reading program. An effect score of 1.84 indicates that FEP students continued to make more than one year’s growth in fifth grade. For year five, the FEP effect score continues to decrease from year four (see table 13).

Table 13

Effect Size Analysis Applying a Paired T-Test to Measure Growth Between Pre and Post WPM Scores for SY 2011 (Year 5)

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	99% Confidence Interval of the Difference		t	df	Sig. 2-tailed
				Lower	Upper			
LEP Group	34.800	16.398	3.667	27.125	42.475	9.491	19	.000
FEP Group	27.650	15.013	3.357	20.624	34.676	8.236	19	.000

In the fifth and final year of the study, students in both groups continue to demonstrate significant growth in reading scores. After consistent growth over four years of the study, the LEP group demonstrates a decrease in growth. The FEP group makes a slight increase from year four. Students in the LEP group make a larger average gain in effect size by a difference of .28 indicating less than one year’s growth over their FEP peers making it the least significant difference in effect size between both groups. One thing to note is that students in the LEP program are most likely to exit the 4-hour ELD block by fifth grade, which could explain why effect scores are more similar.

Summary of Analysis

The preponderance of evidence demonstrates that both the LEP and FEP groups made significant gains while in their perspective programs during each year of the study. However, there was no significant difference in results between the LEP and FEP group for each measure. This leads to conclude that the success of the students in this program was most likely due to factors other than the specific programming. In other words, the results demonstrate that students with the same reading abilities will progress at a similar rate regardless of the program they are enrolled in.

What can be concluded from the data is that the LEP group maintained a longer pattern for growth over the study. The FEP group had a more erratic path in growth. Although there is not enough information to determine causality, a hypothesis can be made based on the data presented (See figure 1). The 4-hour ELD block is not a scripted program but it does have strict protocols to follow including meeting specific language objectives. The structured protocols could create a more consistent professional development for teachers who then provide a more consistent instructional program. It could also be possible that students move from classroom to classroom as a cohort. This means that as students move up in grades, they have similar ability levels that allowed the teaching to focus on closer ability levels and have less need for differentiation. The results are that students in the 4-hour ELD program have more opportunities to build on the knowledge and growth from the year prior. A regular reading classroom would have a wider range of reading abilities.

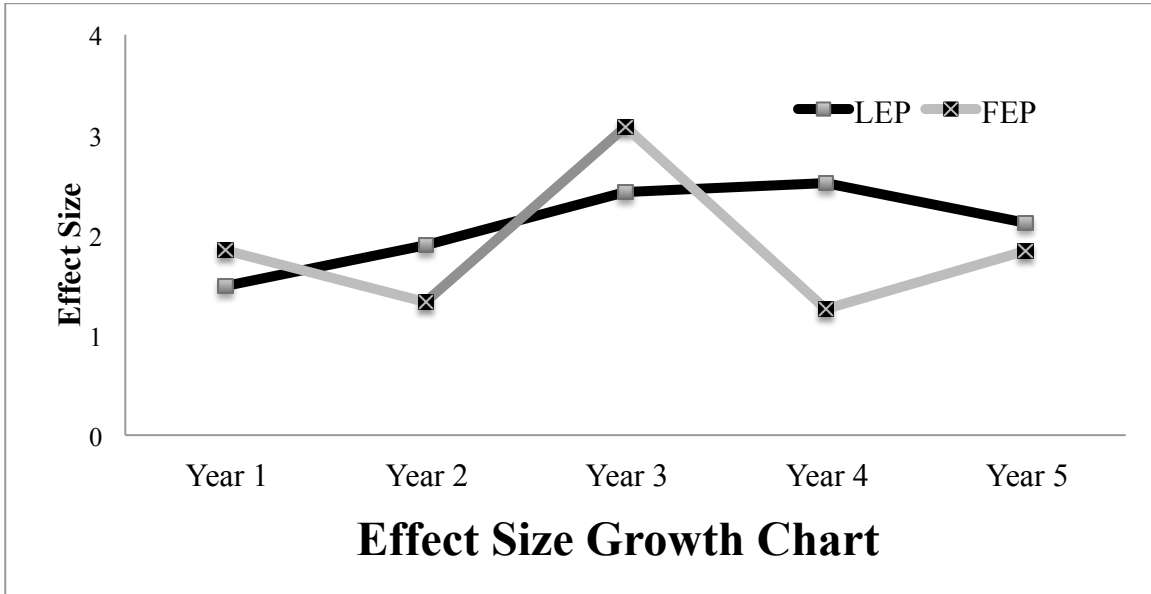


Figure 1. Effect Size for LEP and FEP Groups.

AIMS Results

The percent passing for the AIMS test in reading will be graphed to compare the academic achievement between the LEP and FEP groups. The AIMS achievement scores are used by the state to determine if a school or grade level has achieved adequate yearly progress (AYP). The state assumes that a school or grade level with 80% passing has adequately made a year’s growth. Therefore, 80% passing will also be used to interpret the results presented. The percent passing on AIMS reading will be analyzed for years three through four of the study, the only years the AIMS reading were taken by the participants.

AIMS reading results for year three. In year three of the study, eighteen out of twenty students from the LEP passed the AIMS test in reading or approximately 90% of the LEP group. Two students in the LEP group failed the AIMS test in reading, which is approximately 10% of the LEP group. Fourteen out of twenty students in the FEP group

passed the AIMS test in reading or approximately 70% of the FEP group. Six students in the FEP failed, which is approximately 30% failing. When measured to all participants in the study, approximately 45% of the LEP group passed the AIMS test for reading compared to approximately 35% of students in the FEP group that passed it (figure 2).

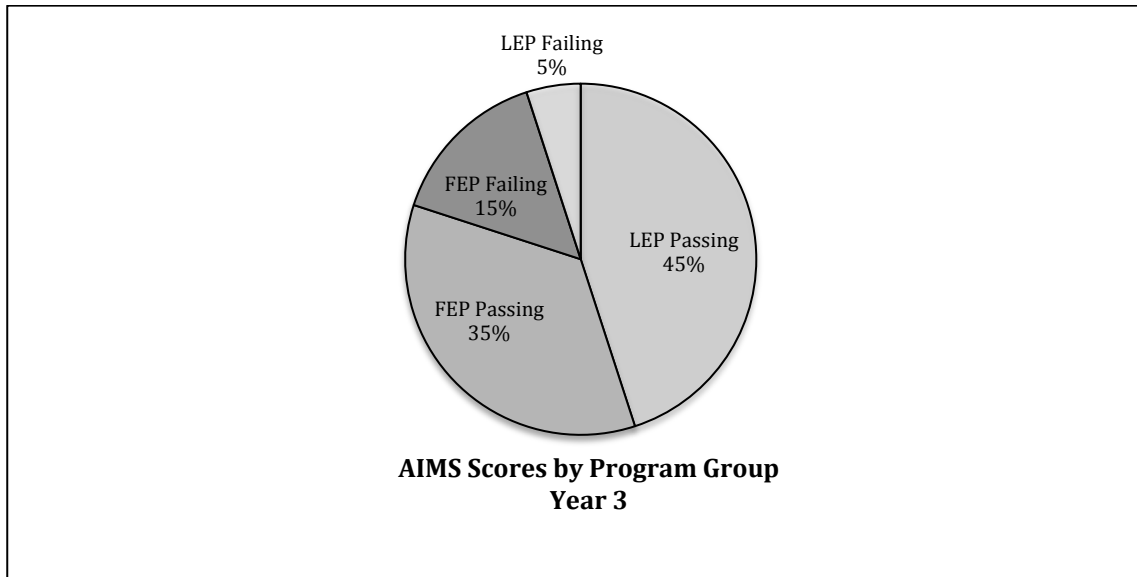


Figure 2. Pass/Fail Percentages for Both Groups (Year 3).

The data also indicates that 5% of students in the LEP group passed the AIMS test for reading with a score of “Exceed” while 40% of the group passed the AIMS test for reading with a score of “Meet”. 5% of the students in the FEP group passed the AIMS test for reading with a score of “Exceeds” while 30% of the students in the FEP group passed the AIMS test for reading with a score of “Meets”.

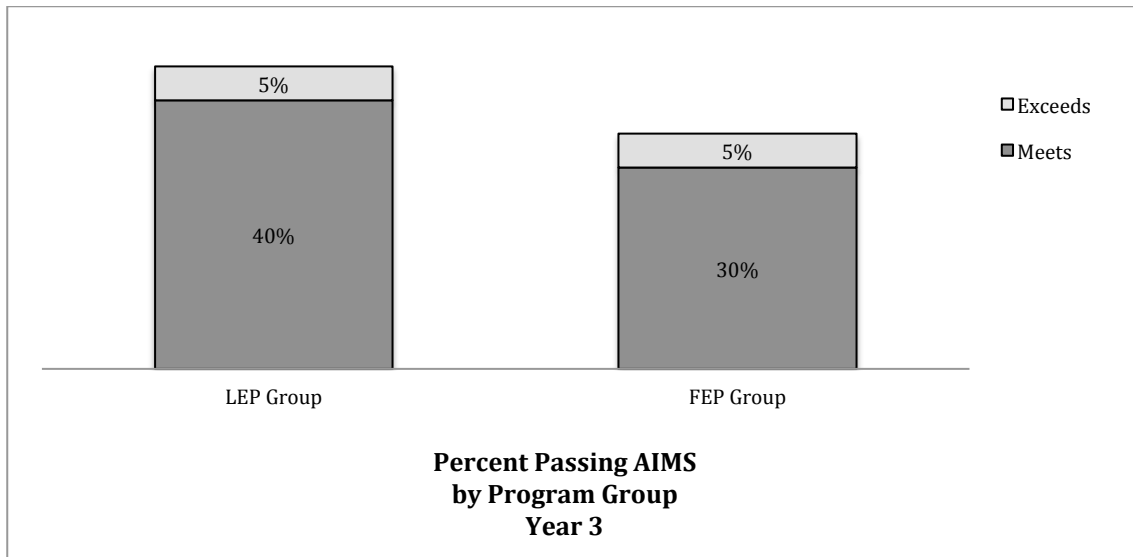


Figure 3. Percent Passing (Year 3).

According to the data shown in figure 3, The LEP group made adequate yearly progress in the third year of the study. The FEP group did not make adequately year progress. The FEP group was short by 10% or approximately four students passing. More students in the LEP group passed the AIMS test for reading than students in the FEP group. The LEP group had 5% more students passing than the FEP group. Both groups had 5% of students pass the AIMS reading test with a score of “Exceed”. Results indicate the LEP group achieved better than the FEP group but not significantly better.

AIMS reading results for year four. In year four of the study, eighteen out of twenty students from the LEP passed the AIMS test in reading or approximately 90% of the LEP group. Two students in the LEP group failed the AIMS test in reading, which is approximately 10% of the LEP group. Fourteen out of twenty students in the FEP group passed the AIMS test in reading or approximately 70% of the FEP group. Six students in the FEP failed, which is approximately 30% failing. When measured to all participants in the study, approximately 45% of the LEP group passed the AIMS test for reading

compared to approximately 35% of students in the FEP group that passed it (figure 4).

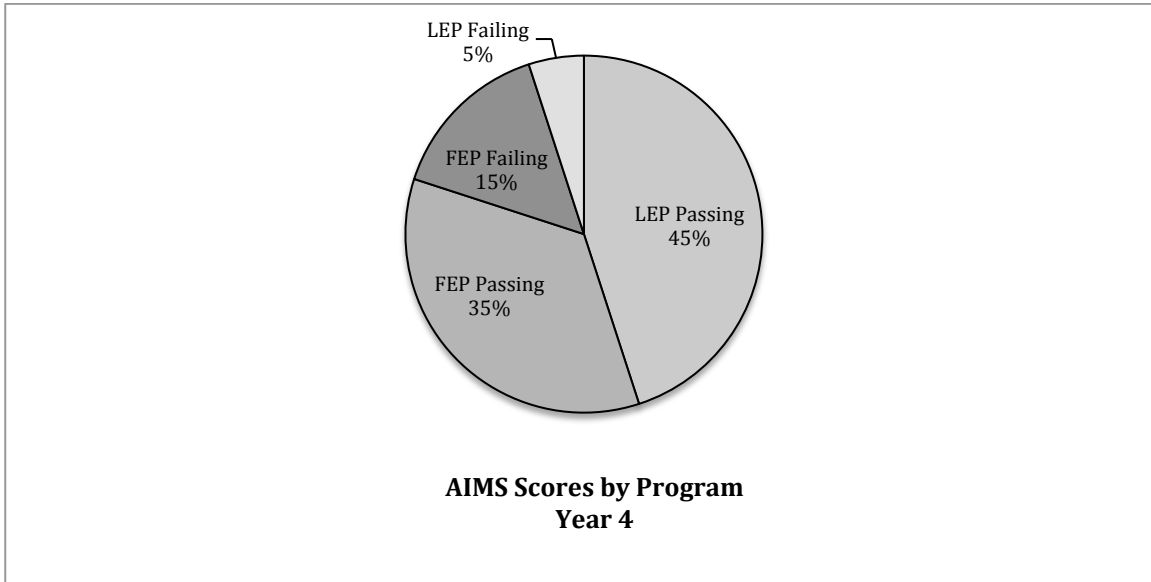


Figure 4. Pass/Fail Percentages for Both Groups (Year 4).

The data also indicates that 5% of students in the LEP group passed the AIMS test for reading with a score of “Exceed” while 40% of the group passed the AIMS test for reading with a score of “Meet”. 8% of the students in the FEP group passed the AIMS test for reading with a score of “Exceeds” while 28% of the students in the FEP group passed the AIMS test for reading with a score of “Meets”.

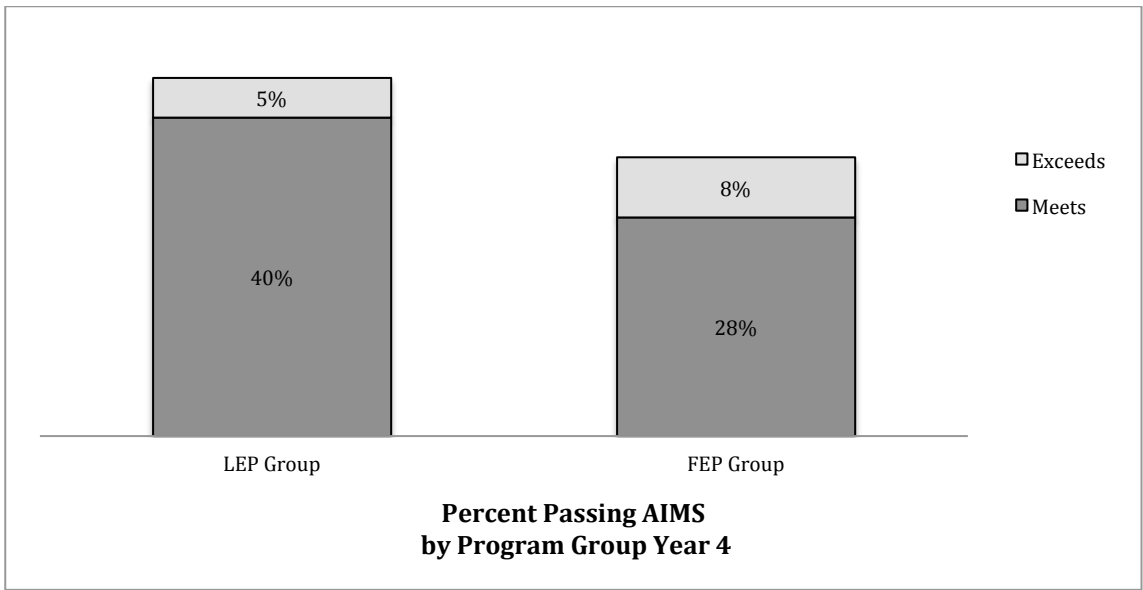


Figure 5. Percent Passing (Year 4).

The results for year four, shown in figure 5, were similar to those in year three indicating neither program had an effect on achievement. However, the FEP group increased the percent of students with a passing score of “Exceeds” from 5% to 8%. The LEP group had the same results. According to the data, the LEP group made adequate yearly progress in year four of the study. The FEP group did not make adequate yearly progress by 10%. The FEP group increased the percent scoring “Exceeds” by 3%. The results indicate that the LEP group achieved higher than the FEP group, however the FEP group had more students improve their AIMS score to “Exceeds”.

AIMS reading results for year five. In year five of the study, seventeen out of twenty students from the LEP passed the AIMS test in reading or approximately 85% of the LEP group. Three students in the LEP group failed the AIMS test in reading, which is approximately 7% of the LEP group. Sixteen out of twenty students in the FEP group passed the AIMS test in reading or approximately 70% of the FEP group. Four students

in the FEP failed, which is approximately 20% failing. When measured to all participants in the study, approximately 43% of the LEP group passed the AIMS test for reading compared to approximately 40% of students in the FEP group that passed it (figure 6).

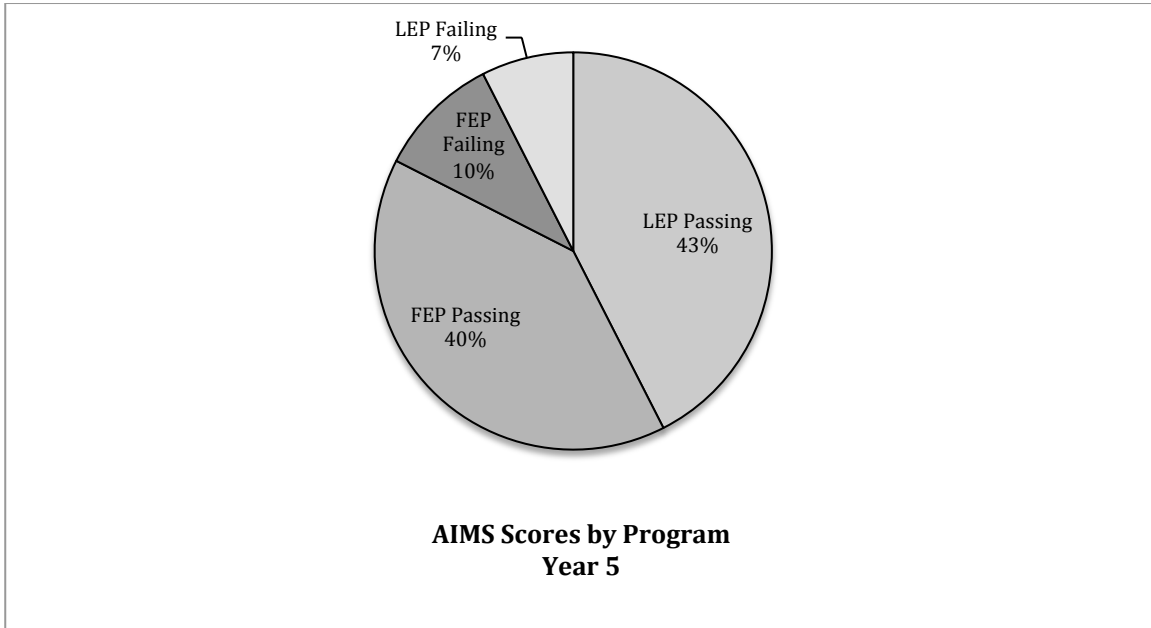


Figure 6. Pass/Fail Percentages for Both Groups (Year 5).

The data also indicates that 8% of students in the LEP group passed the AIMS test for reading with a score of “Exceed” while 35% of the group passed the AIMS test for reading with a score of “Meet”. 3% of the students in the FEP group passed the AIMS test for reading with a score of “Exceeds” while 38% of the students in the FEP group passed the AIMS test for reading with a score of “Meets”.

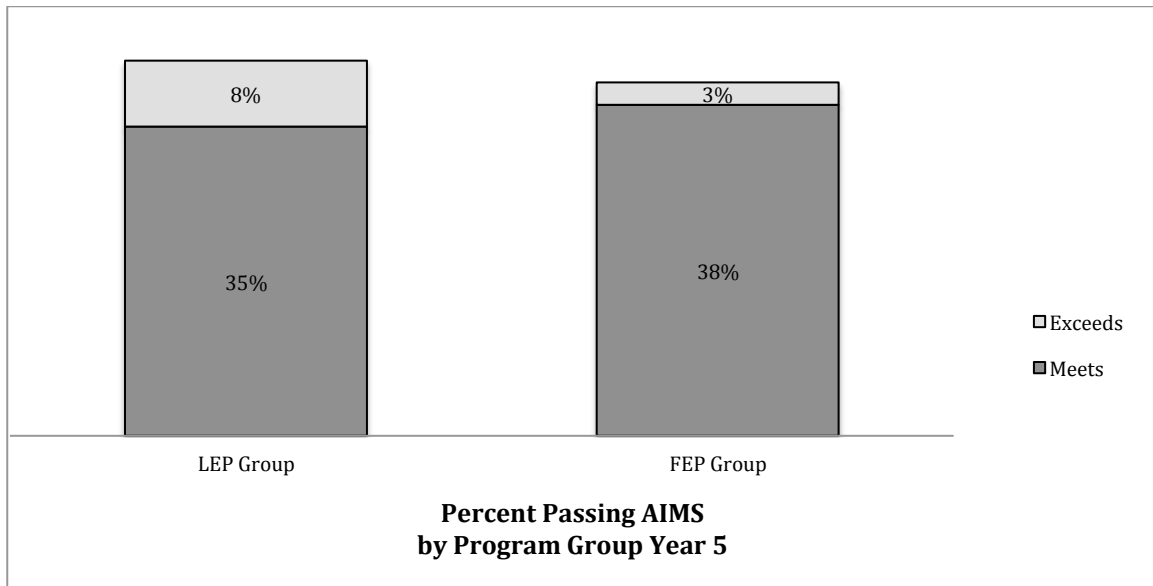


Figure 7. Percent Passing (Year 5).

According to the data in figure 7, The LEP and FEP group made adequate yearly progress in the fifth year of the study. More students in the LEP group passed the AIMS test for reading than students in the FEP group. The LEP group had 3% more students passing than the FEP group. The LEP group had 8% of students with a passing score of “Exceed” while the FEP group had 3% of students with a passing score of “Exceeds. The results indicate both groups increased achievement scores in reading by the fifth year of the study to achieve adequate yearly progress. There is a difference regarding the increase in percent of students receiving a passing score of “Exceeds” for the LEP group and the decrease in percent of students receiving the same score. This suggests that the LEP group seemed to consistently increase achievement scores over time while the FEP group was less consistent in increasing or maintaining achievement scores.

Conclusion of the Chapter

The results of three measures were taken in the study to determine the impact of

the 4-hour ELD block. The results were examined and a conclusion was reached that the 4-hour ELD block did not make a significant impact when compared to similar students that were not in the program. A summary of the findings is presented in this chapter.

The student t-test reported that when you compare the benchmark reading scores for students in the LEP group to the FEP group, there was not significant difference in scores. This can be interpreted in two ways. The first assumes that both groups had students with similar reading scores when starting the study. Since no significant difference was found, than it can be argued that student in the LEP group, or those in the 4-hour ELD block, did not have a statistically significant benefit by being in the program. The second assumes that the LEP group has limited English skills and the 4-hour ELD block was the reason they were able to score similar to the FEP group. In other words, without the intervention, the LEP group would show scores that were significantly lower than the FEP group. However, a limitation to the study is that the LEP group took the AzELLA tests and did not pass in that initial year despite having similar reading abilities as their FEP counterparts. Furthermore, the FEP group did not take the AzELLA test. Therefore, we do not know if students in the FEP group would have scored better on the AzELLA which means it is not conclusive that the LEP group were academically behind the FEP group and “caught up” because of the intervention. Because the LEP group were academically similar in reading at the start of the study, it seems more plausible that the LEP group would have achieved similar growth rates as the FEP group if they were not enrolled in the 4-hour ELD block.

Examining the results of AIMS scores supports the hypothesis that students in the LEP group did just as well academically as their FEP counterparts. After five years, the

percent of students in the LEP group that passed the reading portion of the AIMS test were similar to the percent of students in the FEP group passing the same test. Both groups had more than 60% of the students in each group passing the reading portion of the AIMS test meaning both groups achieved adequate yearly progress. However, for the AIMS measure, the actual percent passing should not be considered an indicator for impact because this is a small test sample where students in each group were of similar ability. What should be considered is that between two similar groups, the percentage of students passing was similar concluding there was not a significant difference in outcomes for either program.

This leads to the examination of the effect size. The LEP group reported more than two years growth for each year in the study. This could indicate that the 4-hour ELD block had a positive impact especially when you conclude that the LEP group made adequate yearly progress. However, the FEP group reported similar data, which aligns more to the argument that the LEP group did not do significantly better or worse than the FEP group. What does stand out in the effect size measure is when you graph the effect sizes for each group and for each year of the study on a line graph. The line for the FEP group oscillates from year to year thereby creating peaks and valleys while maintaining an overall positive trend line. The line for the LEP group trends more of a curve with smaller peaks and valleys as it follows nearly the same positive trend line. Both lines are consistent with the growth measured in effect size and AIMS percentages. However, the LEP line indicates that growth was more consistent for the students in the LEP group than for students in the FEP group. The pattern suggests that the 4-hour ELD block was more substantive in supporting students from falling behind in achievement than the

regular curriculum.

It can be concluded then that the 4-hour ELD block program had the same impact on reading achievement as the regular curriculum. These findings are consistent with research on state and national data that looked at growth for ELL students. It can also be concluded that students in the 4-hour ELD block made substantial growth from year to year and on the state AIMS which seems consistent with the school level research on the SEI model. However, the growth by LEP students was equivalent to the same growth made by FEP students in the regular curriculum during the same time period. What is noticeable is that growth for the LEP students was more consistent than the growth for the FEP students. A pattern that is consistent with an intervention, which supports students' learning.

Chapter 5: Conclusion And Discussion

The research examined reading achievement scores between LEP and non-ELL students in order to answer the following questions. What is the impact of the 4-hour ELD Block in Arizona elementary schools with LEP students that make up more than 50% of the population? What is the rate of progress made by students in the 4-hour ELD Block versus their English proficient peers in the general curriculum? What are possible explanations for differences in achievement between students in the 4-Hour ELD Block and their English proficient peers? The responses are provided in this chapter.

Impact of the 4-hour ELD block

In regards to the first question on the impact of the 4-hour ELD block, the results conclude that the LEP group in the study made significant gains in reading. However, the results also concluded that the FEP group made similar gains in reading as well. It implies that the 4-hour ELD block made no impact as a reading program. It does suggest that the gains made are indicative of an effective program, which concurs with the evaluations made in the SEI studies. But the results indicate that the program's effectiveness is more likely due to how it was taught, not that it was a significantly better program. Similar results for both programs indicate that the LEP students would have done just as well had they been in the regular curriculum. In this context, reading achievement results for the LEP group can be predicted by the results of the reading achievement results for the FEP group. In other words, if most of the FEP population in a school demonstrates low reading achievement, then the LEP population would as well. It may explain why the current research that used AIMS found that the reading achievement gaps the ELL group are remaining consistent.

The findings are significant considering students are identified and placed in the 4-hour ELD block based largely on a parent survey and the results of a language test. The student's reading ability is not taken into consideration. If the 4-hour ELD program is proven to be as effective as the regular curriculum, than it is possible that ELL students with higher reading abilities are needlessly being placed in the program. A second consideration is for ELL students who demonstrate reading deficiencies as opposed to language deficiencies. In essence, ELL students who are struggling readers are being inappropriately placed in a language intervention instead of a reading intervention to address their needs. This would be more significant in schools with more than 50% of the population being ELL, which would place more students in the language intervention, and essentially creating the aforementioned "Ethnic Rooms". Furthermore, resources would be expended that could otherwise be used more prudently.

In regards to the second question on comparing the rate of progress, the LEP groups demonstrated similar rates of progress as the FEP group in reading achievement. The lack of any significant difference, in the researcher's opinion, indicates that there is little justification for having two programs that essentially produce the same result. Had differences been presented, than the differences could have been examined and effective causes replicated. This is important knowing that there will always be ELL students that need English language instruction. What is implied is that whether it be the 4-hour ELD block or another language acquisition program, it could be adjusted to increase academic achievement results.

Finally, in regards to the third question on identifying any difference, the results demonstrated there was no significant difference in reading achievement scores.

However a pattern emerged when comparing growth data between the two groups. The LEP group demonstrated a more consistent projection in growth having minimal peaks and valleys in the data. The FEP group demonstrated larger shifts between years indicating larger growth and regression from year to year. This indicates that the LEP demonstrated less regression year to year than the FEP group, which may be a benefit of being in the 4-hour ELD block.

In the researchers opinion, the consistency of growth may have resulted from taking a proactive approach to intervention as opposed to a reactive approach to intervention. Students in the LEP group are considered to have a language deficit and therefore teachers take extra precautions when working with these students. Teachers take time to assess progress, provide extra support, and apply strategies that promote learning for struggling readers. Students in the FEP group are considered “normal” and therefore teachers are most likely to teach to the average student. Students in the FEP group may struggle in reading but it would not manifest until formal assessments are given. By then, struggling readers in the FEP group may have lagged behind. The growth for the FEP group indicates that struggling readers did get the intervention they needed, but they had to fail first. The consistency in growth for the LEP group indicates that struggling readers in the group had more opportunities and faster support to maintain their reading development.

The 4-hour ELD block and meritocracy

Meritocracy is the political philosophy that believes all free citizens of a democracy should have an equal opportunity to achieve their aspirations and financial goals. To that end, rewards are allocated to individuals in society based on their

performance, achievement, and merit and not on factors like family socio-economic status or gender. Meritocracy has played out in public education, as the places were children from any background could use education to determine their future. However the Coleman report demonstrated that that not all children have an equal start. Since then, the United States have continued to enact laws and provide funding to ensure all students have equal access to education and the opportunity to succeed. Through this lens, the 4-hour ELD block is a result of trying to provide Hispanic ELL students access to an equal education. The findings of this study are not meant to get rid of the 4-hour ELD block program, but to ensure education is providing the right one.

Final Remarks

The findings of this study are not conclusive, but it adds to the discussion regarding the use of the 4-hour ELD block. I believe that it is agreed that English language learners need a language acquisition program, and I believe that the search of an effective program is warranted. This study does not address how the 4-hour ELD block is effective as a language program. The results of this study have demonstrated that it can be effective as an intervention and it is a model of proactive instruction. There are some recommendations I would like to share for consideration.

The first recommendation is that the proficiency of a student's first language and the students reading ability should be used as part of the identification process. What the research indicates is that ELL children who are not proficient in their first language will struggle to learn a second. Placing them in the 4-hour ELD block would not be appropriate without other language and reading supports. Conversely, should an ELL student demonstrate strong abilities, that student should be allowed to receive instruction

in the regular curriculum if it means they will have access to the resources and activities that will be limited to her or him if they were in the 4-hour ELD block.

The second recommendation is that students in the primary grades should not be enrolled into the 4-hour ELD block until it can be established that a student is proficient in their first language. Students in kindergarten to third grade should receive reading and language development instruction especially in schools with a high ELL and poor population. The research demonstrates that students of poor communities are most likely to have deficits in language prior to starting kindergarten.

REFERENCES

- Arizona State Senate. (2013). Arizona State Senate Issue Brief: Flores v. Arizona. Phoenix: Arizona Senate Research Staff.
- Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Council for Exceptional Children, 71*(3), pp. 283 – 300.
- Atwill, K., Blanchard, J., Christie, J., Gorin, J., & Garcia, H. (2010). English language learners: implications of limited vocabulary for cross-language transfer of phonemic awareness with kindergartners. *Journal of Hispanic Higher Education, 9*(2), pp. 104 – 129. doi:10/1177/1538192708330431
- Barnes, Carol A. (2002). *Standards Reform in High Poverty Schools: Managing Conflict and Building Capacity*. NY: Teacher's College Press.
- Bernstein, B. (1971). *Class, Codes, and Control. Volume 1: Theoretical Studies towards a Sociology of Language*. Great Britain: Routledge & Kegan Paul Ltd.
- Bourdieu, P. (1977). The economics of linguistic exchanges. *Social Science Information, 16*(6). pp. 645 – 668. doi:10.1177/053901847701600601
- Bourdieu, P. (1991). *Language and Symbol of Power*. Cambridge: Policy Press in association with Blackwell.
- Burney, V & Beilke J. (2008). The constraints of poverty on high achievement. *Journal for the Education of the Gifted, 31*, pp. 171 – 197. Retrieved from <http://www.prufrock.com>
- Clark, K. (1999). From primary language instruction to english immersion: how five California districts made the switch. (READ Abstracts Research and Policy Brief. Washington DC: Institute for Research in English Acquisition and Development (READ). Retrieved from <http://read-institute.org>
- Clark, K. (2009). The case for structured english immersion. *Educational Leadership, 66*(7), 42-46.
- Cohen, J., (2007). A case study of a high school English-language learner and his reading. *Journal of Adolescent & Adult Literacy, 51*, pp. 164 -175. Retrieved from <http://www.jstor.org/stable/40021845>

- Colman, J. S., (1966). *Equality of Educational Opportunity* (Report number OE-3800). Washington DC: US Department of Health, Education & Welfare – Office of Education.
- Cummins, J. (1984). *Bilingual Education and Special Education: Issues in Assessment and Pedagogy*. San Diego: College Hill.
- Cummins, J. (2000). *Language, Power, and Pedagogy: Bilingual Children in the Crossfire*. Cleveland: Multilingual Matters.
- Echevarria, J., Powers, K., & Short, D. (2006). School reform and standards-based education: a model for English-language learners. *The Journal of Educational Research, 99*(4), pp. 195 – 210.
- Ellis, R. (1984). *Classroom Second Language Development*. Oxford: Pergamon.
- Ellis, R. (1991). *Second Language Acquisition and Language Pedagogy*. Clevedon: Multilingual Matters.
- Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.
- Florez, I. R. (2010). *Do the AZELLA Cut Scores Meet the Standards? A Validation Review of the Arizona English Language Learner Assessment*. Los Angeles: The Civil Rights Project Proyecto Derechos Civiles. Retrieved from <http://civilrightsproject.ucla.edu/Arizona>
- Fry, R. (2008). *The Role of Schools in the English Language Learner Achievement Gap*. (Report). Washington DC: Pew Hispanic Center. Retrieved from <http://www.pewhispanic.org>
- Gandara, P., & Orfield, G. (2010). *A Return to the "Mexican Room": The Segregation of Arizona's English Learners*. Los Angeles: The Civil Rights Project Proyecto Derechos Civiles. Retrieved from <http://www.civilrightsproject.ucla.edu/Arizona>
- Garcia, E. E., Lawton, K., & Diniz de Figueiredo, E. H. (2010). *The Education of English Language Learners in Arizona: Persisting Achievement Gaps in a Restrictive Language Policy Climate*. Los Angeles. Los Angeles: The Civil Rights Project Proyecto Derechos Civiles. Retrieved from <http://www.civilrightsproject.ucla.edu/Arizona>
- Garcia, E. E., Lawton, K., Diniz de Figueiredo, E. H. (2012). The education of English language learners in Arizona: a history of underachievement. *Teachers College Record, 114*, pp. 1-18. Retrieved from <http://www.tcrecord.org>

- Goldenberg, C., & Rutherford Quach, S. (2010). *The Arizona Home Language Survey: The Identification of Students for ELL Services*. Los Angeles: Civil Rights Project Project Derechos Civiles. Retrieved from <http://www.civilrightsproject.ucla.edu/Arizona>
- Guerrero, M. D. (2004). Acquiring academic language in one year: an unlikely proposition for English language learners. *Urban Education, 39*(2), pp.172-199. doi:10.1177/0042085903260915
- Hart, B. & Risley, T. R. (2003). The early catastrophe: the 30-million word gap by age 3. *American Educator, 27*, Pp. 4-9.
- Hattie, J. (2009). *Visible Learning: A Synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge
- Jimenez- Castellanos, O., Combs, M. C., Martinez, D., & Gomez, L. (2013). *English Language Learners: What's at Stake for Arizona*. Arizona: ASU Morrison Institute for Public Policy. Retrieved from arizonaindicators.org.
- Kelly, P., Gomez-Bellenge, F., Chen, J., & Schulz, M. (2008). Learner outcomes for English language learners low readers in an early intervention. *TESOL Quarterly, 42*(2), pp. 235 – 260.
- Klinger, J. K., Artiles, A. J., Kozleski, E., Harry, B., Zion, S., Tate, W., Duran, G. Z., & Riley, D. (2005). Addressing the disproportionate representation of culturally and linguistic diverse students in special education through culturally responsive educational systems. *Education Policy Analysis Archives, 13*(38). pp. 1-43. Retrieved from <http://epaa.asu.edu/epaa/v13n38>
- Klinger, J. & Artiles, A. J., (2006). English language learners struggling to learn to read: emergent scholarships on linguistics differences and learning disabilities. *Journal of Learning Disabilities, 39*(5), pp. 386 – 389. doi:10.1177/00222194060390050101
- Klinger, J., Artiles, A. J., & Mendez Barletta, L. (2006). English language learners who struggle with reading: language acquisition or LD? *Journal of Learning Disabilities, 38*(2), pp. 108 – 128.
- Krashen, S. D. (1987). *Principles and Practice in Second Language Acquisition*. Oxford: Princeton-Hall International
- Krashen, S. D. (1988). *Second Language Acquisition and Second Language Learning*. Oxford: Princeton-Hall International

- Lacour, M. & Tissington, L. D. (2011). The effects of poverty on academic achievement. *Educational Research and Reviews*, 6(7) pp. 522-527. Retrieved from <http://www.academicjournals.org/ERR>
- Lillie, K. E., Markos, A., Arias, M. B., & Wiley, T. G. (2012). Separate and not equal: the implementation of structured English immersion in Arizona's classrooms. *Teachers College Record*, 144, pp. 1-33. Retrieved from <http://www.tcrecord.org>
- Lillie, K. E., Markos, A., Estrella, A., Nguyen, T., Trifiro, A., Aria, B. M., et al. (2010). *Policy in Practice: The Implementation of Structured English Immersion in Arizona*. Los Angeles: Civil Rights Project Proyecto Derechos Civiles. Retrieved from <http://www.civilrightsproject.ucla.edu/Arizona>
- Lopez, M. H. & Velasco, G. (2011). *The Toll of the Great Recession: Childhood Poverty Among Hispanics Sets Record, Leads Nation*. (Report). Washington DC: Pew Hispanic Center. Retrieved from <http://www.pewhispanic.org>
- Martinez-Wenzel, M., Perez, K., & Gandara, P. (2010). *Is Arizona's Approach to Educating Its ELS Superior to Other Forms of Instruction?* *Teachers College Record*, 114, pp. 1-32. Retrieved from <http://www.tcrecord.org>
- MacSwan, J. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Services*, 22:(1), pp. 3-45.
- MacSwan, J. & Rolstad, K. (2006). How language proficiency tests mislead us about ability. *Teachers College Record*, 108(11), pp. 2304-2328. Retrieved from <http://www.tcrecord.org>
- National Academies of Science, Engineering, and Medicine (U.S.) (2017). *Promoting the Educational Success of Children and Youth Learning English: Promising Futures*. Washington DC: The National Academies Press. [doi:10.17226/24677](https://doi.org/10.17226/24677).
- OELAS, English Language Proficiency Standards. (2019). Retrieved from <http://www.azed.gov/oelas/elps/>
- OELAS, Monitoring. (2019). Retrieved from <http://www.azed.gov/oelas/monitoring>
- Pacheco, M. (2010). English Language learners' reading achievement: dialectical relationships between policy and practices in meaning-making opportunities. *Reading Research Quarterly*, 45(3), pp. 292-317. doi:10.1598/RRQ.45.3.2
- Rios-Aguilar, C., Gonzalez-Canche, M., & Moll, C. L. (2010). A Study of Arizona's Teacher of English Language Learners. *Teacher's College Record*, 114, pp. 1-33. Retrieved from <http://www.tcrecord.org>

- Rios-Aguilar, C., Gonzalez-Canche, M., & Moll, L. C. (2010). *Implmenting Structured English Immersion in Arizona: Benefits, Costs, Challanges, and Opportunities*. Los Angeles: The Civil Rights Project Proyecto Derechos Civiles. Retrieved from <http://www.civilrightsproject.ucla.edu/Arizona>
- Rios-Aguilar, C., Gonzalez-Canche, M., & Sabetghadam, S., (2012). Evaluating the impact of restrictive language policies: the Arizona 4-hour English language development block. *Language Policy: Springer, 11*, pp. 47 -80. doi:10.1077/s10993-011-9226-3
- Rios-Aguilar, C., Gandara, P., (2012). Horne v. Flores and the future of language policy. *Teacher College Record, 114*. pp. 1-13. Retrieved from <http://www.tcrecord.org>
- Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: a meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*, pp. 572 – 594. doi:10.1177/0895904805278067
- Rubenstein-Avila, E., (2004). Conversing with Miguel: an adolescent English language learner struggling with later literacy development. *Journal of Adolscent & Adult Literacy, 47:4*, pp. 291 – 301. Retrieved from <https://www.jstor.org/stble/40014775>
- Slavin, E. R., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75 (2)*, pp. 247-284. Retrieved from <http://www.jstor.org/stable/3516050>
- Stanovich, Keith (1986). Matthew Effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21(4)*, 360 – 406.
- Sullivan, A. L., (2011). Disproportionality in special education identification and placement of English language learners. *Council for Exceptional Children, 152(3)*, Spring, pp. 317 – 334.
- Taguchi, N., (2007). Development of speed and accuracy in pragmatic comprehension in English as a foreign language. *TESOL Quarterly, 4(2)*, pp. 313 – 338. Retrieved from <http://www.jstor.org/stable/40264355>
- US Department of Education. (2011). Consolidated State Performance Report: Parts I and II. For State Formula Grant Programs for Reporting on Shool Year 2009-10 Arizona. Washington DC:Washinton DC.

US Department of Education. (2011). *National Center for Education Statistics*. (U. D. Education, Producer) Retrieved January 4, 2014, from National Assessment of Educational Progress (NAEP):
nces.ed.gov/nationsreportcard/pdf/studies/2011485.pdf