AN EVALUATION OF THE EFFECTS OF A CHANGE IN THE SCREENING METHODOLOGY ON THE READING ACHIEVEMENT OF ELEMENTARY SCHOOL ELLS USING SHORT INTERRUPTED TIME SERIES ANALYSIS

A thesis presented
by
Victoria Lazareva-Stanishcheva

to
The School of Education

In partial fulfillment of the requirements for the degree of
Doctor of Education

in the field of
Education

College of Professional Studies
Northeastern University
Boston, Massachusetts
December, 2014
Abstract

While the population of English language learners (ELLs) in US schools has been increasing, many ELLs demonstrate a lower level of reading skills compared to other students. Educators have been exploring various options to address this issue. One option is to provide initial reading instruction in the primary (L1) language. A considerable body of research suggests that L1 reading instruction can be beneficial for the English reading of ELLs. However, little attention has been paid to the role of L1 proficiency in enhancing English reading of ELLs. This quasi-experimental study attempted to fill this gap by examining the effects of an intervention represented by a change in the screening methodology for placing Spanish-speaking ELLs either in L1 or English-only reading instruction in grades K-3. Using short interrupted time series, the study examined reading achievement of ELLs before and after the intervention as shown by the scores on the third grade standardized reading achievement test. The results of segmented regression revealed that the intervention had a positive effect on the post-intervention trend in the means of scores with respect to all ELLs. The model with an interaction term demonstrated a statistically significant positive difference in the trends of the means of scores between the group instructed in L1 and the group instructed in English only during the post-intervention period. No statistically significant difference in the means of scores between the pre- and post-intervention periods was observed. The study findings contribute to the research supporting the premise about the positive role of L1 reading instruction for ELLs, draw attention to the importance of L1 proficiency of ELLs, and promote the use of interrupted time series in education research.

Key words: English language learner (ELL), language proficiency, language of reading instruction, first (native, primary) language reading instruction, bilingual reading program, transitional bilingual education (TBE), interrupted time series, segmented regression.
Dedication

This work is dedicated to my grandparents Afanasiy Antonovich Bogatyrev and Alexandra Alexandrovna Bogatyreva, my husband John F. Douglass, Ed.D., and my son Vladimir S. Lazarev-Stanishchev with love.
Acknowledgments

Many people helped me to bring this project to conclusion. I would like to start with acknowledging my initial advisor Dr. Anita Kite and second reader Dr. Elizabeth Bennett. Dr. Kite helped to modify the research design by suggesting that I should use interrupted time series analysis, an effective approach unfortunately not so commonly used in education research. With Dr. Kite’s guidance, I was able to successfully prepare and defend my thesis proposal. Dr. Bennett provided a number of insightful comments that helped to fine-tune my proposal.

Next, I would like to acknowledge the people who helped me to complete my project. Dr. Margaret Gorman seamlessly took over from Dr. Kite and efficiently guided me through the second stage culminating with the thesis defense. Dr. Al McCready, as a second reader, and Dr. Chris Johnson, as an outside examiner, were invaluable in providing their critical feedback. I would like to sincerely thank my advisors, second readers, and the outside examiner!

Further, my heartfelt thanks go to Michael Frisby, a statistical consultant with the Indiana Statistical Consulting Center (ISCC) at Indiana University, Bloomington, IN. He helped me with the statistical aspect of my project, and I very much appreciate his expertise and the patience he demonstrated when working with me.

I would also like to extend my thanks to the ELL Coordinator at the school district where I conducted my research for helping me with the data collection.

Last but not least, this endeavor would not have even taken place without continuous and unwavering encouragement, support, and advice given by my husband John F. Douglass, Ed.D., who was always by my side during this fascinating and amazing journey. There are no words that could fully express my gratitude to him!
Table of Contents

Abstract .......................................................................................................................... 2
Dedication ..................................................................................................................... 3
Acknowledgments ....................................................................................................... 4
List of Tables ............................................................................................................... 9
List of Figures ............................................................................................................. 10
Preface ......................................................................................................................... 11

CHAPTER 1: INTRODUCTION ..................................................................................... 14
  Statement of the Problem ....................................................................................... 14
  Background of the Study ....................................................................................... 17
  Theoretical Framework ......................................................................................... 24
    Common underlying proficiency (CUP) model ...................................................... 24
    Linguistic interdependence ................................................................................. 27
    Cognitive academic language proficiency ......................................................... 29
    Threshold hypothesis ......................................................................................... 33
  Summary ................................................................................................................. 35
  Research Questions ............................................................................................... 35
  Significance of the Study ....................................................................................... 36
  Definition of Terms ............................................................................................... 38

CHAPTER 2: REVIEW OF THE LITERATURE ............................................................. 41
  Introduction ............................................................................................................. 41
  Second Language Literacy Development in ELLs ..................................................... 42
    Literacy development in ELLs and native speakers ............................................. 42
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>68</td>
</tr>
<tr>
<td>Procedures</td>
<td>70</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>71</td>
</tr>
<tr>
<td>Data Analysis Plan</td>
<td>78</td>
</tr>
<tr>
<td>Preparation of the data</td>
<td>78</td>
</tr>
<tr>
<td>Choice of statistical technique</td>
<td>79</td>
</tr>
<tr>
<td>Data analysis</td>
<td>81</td>
</tr>
<tr>
<td>Validity, Reliability, and Generalizability</td>
<td>84</td>
</tr>
<tr>
<td>Validity of the research design</td>
<td>85</td>
</tr>
<tr>
<td>Reliability of the research design</td>
<td>88</td>
</tr>
<tr>
<td>Validity of the language proficiency measure</td>
<td>88</td>
</tr>
<tr>
<td>Reliability of the language proficiency measure</td>
<td>89</td>
</tr>
<tr>
<td>Generalizability</td>
<td>90</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>90</td>
</tr>
<tr>
<td>CHAPTER 4: RESULTS</td>
<td>92</td>
</tr>
<tr>
<td>Research Questions</td>
<td>92</td>
</tr>
<tr>
<td>Sample Characteristics</td>
<td>93</td>
</tr>
<tr>
<td>Power Analysis</td>
<td>95</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>96</td>
</tr>
<tr>
<td>Inferential Statistics and Findings (RQ1)</td>
<td>100</td>
</tr>
<tr>
<td>Inferential Statistics and Findings (RQ2)</td>
<td>107</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>116</td>
</tr>
<tr>
<td>CHAPTER 5: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</td>
<td>119</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Restatement of the Problem</td>
<td>119</td>
</tr>
<tr>
<td>Restatement of the Theoretical Rationale</td>
<td>124</td>
</tr>
<tr>
<td>Limitations</td>
<td>125</td>
</tr>
<tr>
<td>Conclusions</td>
<td>129</td>
</tr>
<tr>
<td>Discussion</td>
<td>130</td>
</tr>
<tr>
<td>Theoretical Implications</td>
<td>135</td>
</tr>
<tr>
<td>Practical Implications</td>
<td>138</td>
</tr>
<tr>
<td>Recommendations for Future Research</td>
<td>142</td>
</tr>
<tr>
<td>Final Comments</td>
<td>145</td>
</tr>
<tr>
<td>References</td>
<td>147</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 Flowchart of the Native Language Literacy (NLL) Program ........................................... 24
Table 2 Formation of Pre-And Post-Intervention Student Cohorts .................................................. 65
Table 3 Data Generation Time Points ............................................................................................... 67
Table 4 Sample Structure Summary .................................................................................................. 95
Table 5 Descriptive Statistics (Test score) ......................................................................................... 97
Table 6 Frequency and Relative Percent (Categorical Variables) ...................................................... 98
Table 7 Estimated Marginal Means .................................................................................................... 100
Table 8 Results of Segmented Regression Analysis (Overall Model) ................................................. 104
Table 9 Results of Segmented Regression Analysis (L1 Model) ....................................................... 109
Table 10 Results of Segmented Regression Analysis (L2 Model) ..................................................... 110
Table 11 Segmented Regression Output (Interaction Model) ............................................................. 113
Table 12 Results of Shapiro-Wilk Normality Test. ........................................................................... 116
List of Figures

Figure 1. The separate underlying proficiency (SUP) model of bilingual proficiency. ........................................................................................................ 25

Figure 2. The common underlying proficiency (CUP) model of bilingual proficiency. ........................................................................................................ 26

Figure 3. The “dual-iceberg” representation of bilingual proficiency. ................................................. 27

Figure 4. The “iceberg” representation of language proficiency. ..................................................... 31

Figure 5. Developmental trends of CALP and BICS. ............................................................................. 32

Figure 6. The “dual-iceberg” representation of bilingual proficiency. ................................................. 33

Figure 7. Mean Scores Through Time (Overall Model). ........................................................................ 105

Figure 8. Q-Q Plot (Overall Model). ................................................................................................. 107

Figure 9. Mean Scores Through Time (L1 Model). ............................................................................. 110

Figure 10. Mean Scores Through Time (L2 Model). .......................................................................... 111

Figure 11. Q-Q Plot (L1 Model). ........................................................................................................ 115

Figure 12. Q-Q Plot (L2 Model). ........................................................................................................ 115

Figure 13. Q-Q Plot (Interaction Model). ............................................................................................ 116
An Evaluation of the Effects of a Change in the Screening Methodology on the Reading Achievement of Elementary School ELLs Using Short Term Interrupted Time Series Analysis

Preface

All around the world, there are many students whose first language (L1) they learned to speak at home is different from the language they are expected to use at school (Durgunoğlu, 2002). The school language they need to learn is an additional or second language (L2). Therefore, these students, referred to as “language learners” (LLs), are exposed to at least two languages, one at school and one at home, and have different levels of proficiency in these languages (Durgunoğlu, 2002). The United States is becoming more linguistically and culturally diverse, and this diversity is expected to increase (Domínguez de Ramírez & Shapiro, 2007). In the United States, language learners are referred to as English language learners (ELLs) as upon school entry they begin to acquire English as the language of school instruction. In recent decades, the number of ELLs in the US has been steadily increasing (Nakamoto, Lindsey, & Manis, 2012). From 1995-1996 to 2005-2006, this student population increased by about 60% and reached over five million students (Nakamoto et al., 2012). The fact that almost 80% of ELLs in the US are native Spanish speakers (Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students, 2008) makes the Spanish language the predominant L1 that ELLs acquire before coming to school.

Children who attend schools where the instructional language is different from their home language (L1) face a special challenge: they have to develop academic skills in their L2, a language in which they are not yet fully proficient in (Jongejan, Verhoeven, & Siegel, 2007). Among these academic skills are L2 literacy skills that are critical to the academic achievement of these students (Nakamoto et al., 2012). While many ELLs are successful in developing their
L2 literacy, too many ELLs are not (Slavin & Cheung, 2005). A majority of ELLs “continue to read well below grade level standards and lag behind other students in reading and academic performance” (Baker, Park & Baker, 2012, p. 255). While Spanish-speaking ELLs constitute the largest segment of ELLs in the US (Kindler, 2002), these students “perform below general education standards as compared to non-ELLs” (Domínguez de Ramírez & Shapiro, 2007, p. 795). Reading difficulties in ELLs become a reason for these students to drop out of school (Gottardo, 2002). Due to the fact that ELLs represent the fastest growing population in public schools in the United States (Kindler, 2002) it is important to determine how to ensure the reading success of these students (August & Hakuta, 1997).

According to Slavin and Cheung (2005), reading instruction for ELLs is one of the most important issues in educational policy and practice. At the same time, there is considerable controversy among educators, researchers, and policymakers on how to ensure reading success in ELLs (Slavin & Cheung, 2005). While there are many aspects of reading instruction that need to be addressed with respect to ELLs, the dominant issue is the language of literacy instruction (Slavin & Cheung, 2005). There is some evidence from research that reading programs using students’ first language (L1) can be beneficial for the English language (L2) literacy of ELLs (Branum-Martin, Foorman, Francis, & Mehta, 2010). Bilingual programs where ELLs were entirely or partially instructed in their L1 were widespread in the United States in the 1970s and 1980s (Slavin & Cheung, 2005). Beginning in the 1990s, both state and federal policies curtailed bilingual education in the United States (Slavin & Cheung, 2005). However, about 50% of ELLs in the United States continued to receive reading instruction in their L1 (Goldenberg, 2008). The two most common types of programs are the early transitional model and the paired bilingual model (Baker et al., 2012). In the early transitional
model, learning to read in students’ first language is followed by learning to read in English. In the paired bilingual programs, students learn to read in their native language and in English simultaneously.
CHAPTER 1: INTRODUCTION

Statement of the Problem

Should ELLs be taught to read in English only, from the onset of their school experience or should they be taught to read in their primary language first? This question has been “the topic of greatest controversy and by far the most volatile issue” (Branum-Martin et al., 2010, p. 341) in the ongoing discussion on how to ensure the academic success of an increasing population of ELLs in public schools in the United States. There has been a fierce debate among the proponents of English immersion and the proponents of bilingual programs (Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011).

The English immersion strategy implies that ELLs should learn to read in their L2 (English) from the beginning of their school career (Slavin et al., 2011). While teachers may employ various language development strategies to support ELLs, reading instruction is provided in the English language using English language instructional materials. The main argument of the proponents of English immersion is that L1 reading instruction will only delay the development of English language skills and English language literacy (Slavin et al., 2011). The primary rationale behind this argument is that more time devoted to English reading instruction should enhance learning to read in English. This premise was supported by the studies conducted by Rossell and Baker (1996) and Thomas and Collier (1997).

The main difference between the English immersion and bilingual programs is that bilingual programs allow for a significant and sustained use of L1 as the language of reading instruction. Advocates of bilingual education point out that the use of L1 for academic purposes can help ELLs to avoid “the failure experience” (Slavin et al., 2011) that is likely to occur when students are required to learn to read in English while still acquiring oral English language skills.
Programs where ELLs first learn to read primarily or entirely in their L1 and then gradually transition to English reading in early elementary grades are referred to as transitional bilingual programs. According to Slavin et al. (2011), the time period when ELLs are taught to read in their L1 is the hallmark of transitional bilingual education (TBE). In “early-exit” programs this time period ends in second or third grade, while “late-exit” programs are extended through all elementary grades (Ramirez, Pasta, Yuen, Billings, & Ramey, 1991). Meta-analyses of research on L1 reading instruction suggested that the use of L1 as opposed to English only had a beneficial effect on the English reading of ELLs (Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005). In addition, there is evidence that L1 reading proficiency can be a predictor of L2 reading proficiency (August & Shanahan, 2006). The National Reading Panel on Language-Minority Children and Youth (NRP) concluded that programs providing L1 reading instruction produced a positive effect on the English literacy of ELLs in comparison to the programs providing English-only reading instruction (Francis, Lesaux, & August, 2006). Accordingly, it is feasible to conclude that more recent research has supported the benefits of bilingual education and L1 instruction over English-only instruction (Branum-Martin et al., 2010).

While the NRP findings supported the premise about the positive role of L1 literacy with respect to ELLs, the NRP report stipulated that the level of L1 oral proficiency influenced the initial development of skills related to L2 literacy (August & Shanahan, 2006). This conclusion about the relationship between L1 oral proficiency and L2 literacy development is aligned with the position taken by Durgunoğlu (2002) who stressed the importance of taking into account the level of linguistic proficiency in L1 when examining how L1 literacy skills can support L2 reading achievement. Durgunoğlu (2002) posited that if ELLs had strengths in their L1, these strengths should initially be capitalized on to advance their later literacy development in L2.
Alternatively, according to Durgunoğlu (2002), if children did not have “a strong enough grasp of their L1, possibly because of low home or school support” (p. 201), reading instruction in L2 would be preferable. However, Durgunoğlu, Nagy, and Hancin-Bhatt (1993) pointed out that placement in bilingual programs is usually made on the basis of English (L2) proficiency and staff judgments only. Durgunoğlu et al. (1993) argued that while there appears to be sufficient evidence that L1 literacy skills are transferrable to L2, the specific conditions necessary for this transfer are not clear. Therefore, according to Durgunoğlu et al. (1993), further research was needed to explore how L1 proficiency could affect the transfer of literacy skills to L2. Similar to Durgunoğlu et al. (1993), August and Shanahan (2010) emphasized the need for research studies examining the effect of different approaches to reading instruction with respect to students with different levels of L1 and L2 proficiency. In addition, August and Shanahan (2010) maintained that although it was possible to conclude that L1 reading instruction was generally instrumental in promoting English reading achievement in ELLs, it was still not clear if the benefits of L1 reading instruction are dependent upon the level of L1 proficiency demonstrated by ELLs.

In view of the foregoing, it is possible to conclude that while it has been established that L1 reading instruction can be advantageous for a growing population of ELLs, there is a lack of clarity as to the role L1 proficiency may play in helping ELLs become competent English language readers. Having stated that bilingual programs are beneficial for ELLs and “represent a legitimate policy option” (p. 1979), Cummins (2012), a leading theorist in the field of bilingualism and bilingual education, observed that bilingual instruction by itself cannot be regarded as “a panacea for minority students’ educational difficulties” (p. 1980). Therefore, as suggested by Slavin and Cheung (2005), there is a need for more research that would “illuminate the conditions under which native language reading instruction may be beneficial for developing
English reading skills” (p. 275). In view of the current state of research on the topic of L1 reading instruction for ELLs, there is one issue that has clearly emerged: the relationship between initial L1 oral proficiency and later L2 reading is unclear. As a first step to address this issue, the present study used short interrupted time series analysis to examine if a change in the screening methodology for placing ELLs in L1 or English-only reading instruction influenced later English reading achievement.

**Background of the Study**

The purpose of this quasi-experimental study was to determine whether placement of kindergarten ELLs in either L1 or English-only reading instruction based on their level of L1 proficiency improves scores on a standardized reading achievement test in third grade when compared to the reading scores on the same test of third graders who, in previous years, were assigned to either L1 or English-only reading instruction without regard to L1 proficiency. The study tested the common underlying proficiency model (CUP) (Cummins, 1980a) and the threshold hypothesis (Cummins, 1979) by comparing the English reading achievement (as measured by performance on a standardized reading achievement test) of third grade ELLs who entered school before and after the screening methodology for placing kindergarten students either in L1 reading instruction or English-only reading instruction was changed.

The theory supporting this study will be addressed in detail in Chapter 2. However, for a better understanding of the theoretical rationale it is important to define the construct of bilingualism as a key linguistic factor supporting the concept of bilingual education (Cummins & Swain, 1986). A common belief with respect to the education of ELLs is that if such students come to school with a low level of English language proficiency, they are in need of intensive English language instruction (Cummins & Swain, 1986). Therefore, the use
of students’ L1 to remedy English language deficiency is often viewed by educators and policy makers as counter-productive (Cummins & Swain, 1986). This view is based on the assumption that the maximum school exposure of ELLs to the English language will directly translate into higher academic achievement of these students (Cummins & Swain, 1986). Therefore, Bowen (1977) argued that “the choice of language of instruction in our schools is linguistically irrelevant” (p.116). In contrast to this position, Cummins (1979) maintained that in the discussion of instructional approaches that can be effective for ELLs it is important to recognize linguistic factors, such as bilingualism.

According to Cummins (Cummins & Swain, 1986), bilingualism is a key linguistic factor that needs to be accounted for to understand the relationship between schools and language minority children. This element is “enormously rich and diverse in its complexity” (Cummins & Swain, 1986, p. xiii). At school, bilingual children may experience a unilingual or bilingual environment. When entering school, language proficiency of bilingual children in each language may vary, and each language may be associated with a specific ethnicity or nationality. The definition of bilingualism varies with different researchers and theorists. Macnamara (1967) defined a bilingual as someone who has at least a minimal level of ability in another language. On the contrary, Oeistreicher (1974) posited that bilingualism is conditional on native-like proficiency in a second language. In addition, different levels of bilingualism can be distinguished (Cummins & Swain, 1986). In light of bilingualism as an element of student-school relationship, Cummins (2007) defined bilingual students’ home language as their first language (L1) and the school language as their second language (L2).

The present study used the context of a bilingual program at an elementary school in the Pacific Northwest region of the United States. This bilingual program is set up according to the
early-exit TBE model (Slavin et al., 2011). According to Slavin et al. (2011), in transitional bilingual programs, ELLs are first taught to read primarily or entirely in their L1. Slavin et al. (2011) pointed out that while subjects other than reading can be taught in English, “the hallmark of transitional bilingual education (TBE) is the teaching of reading in the native language for a period of time” (p. 49). In early-exit programs, this period extends through primary grades, and the transition to English-only reading takes place in second or third grade (Ramirez et al., 1991).

The school has chosen to introduce a bilingual program to meet the needs of a considerable and homogenous ELL population. English language learners account for about 51% of the entire student population at the school. All ELLs are native Spanish speakers. Upon entering kindergarten, ELLs are placed either to receive Spanish-language reading instruction or English-only reading instruction. The pull-out model is used for the delivery of Spanish-language reading instruction in kindergarten. Students continue to receive Spanish-language reading instruction through third grade. In first and second grades, Spanish-language reading instruction is delivered in a self-contained classroom. In third grade, students again receive Spanish-language reading instruction on the basis of the pull-out model. The instructional time allocated for Spanish-language reading is one hour in kindergarten and third grade and two hours in first and second grades. Due to the fact that instruction other than reading is provided in the English language, this bilingual program is a transitional bilingual program (Slavin et al., 2011). This bilingual program is also an “early-exit” program because students are exited from the program in third grade (Ramirez et al., 1991).

The bilingual program has been operational at the school for over a decade. The current program format was introduced ten years ago when the school was included in a cohort of elementary schools participating in the Oregon Reading First, a state initiative funded by the
federal government as a part the No Child Left Behind Act of 2001. The goal of this initiative was to help all students become competent readers by the end of third grade. To help achieve this goal, participating schools were provided support through professional development, coaching, and instructional materials. A major element of the Oregon Reading First initiative was the adoption of the school-wide reading program. This adoption implied that all classroom teachers across all grades used the same reading curriculum (Houghton Mifflin Reading Series, 2003) focused on the Five Big Ideas of reading instruction (phonemic awareness, phonics, fluency, vocabulary, and reading comprehension), and used identical instructional strategies (modeling, explicit instruction, meaningful interaction with language, multiple opportunities for practice, corrective feedback, encouragement of student effort, engagement, and success during teacher-led instruction, as well as during independent work). As a part of Oregon Reading First initiative, all classroom teachers in the participating schools received training in curriculum and instructional strategies. Each participating school had to develop a comprehensive school literacy plan. With respect to providing differentiated instruction and supporting ELLs, an important element of the Oregon Reading First initiative was the use of students’ primary language (Spanish) for initial reading instruction.

As it was not feasible to provide initial reading instruction in the Spanish language for all ELLs entering kindergarten, the school that served as a research site for this study had to establish criteria that could be used to place ELLs to receive initial L1 reading instruction versus placing them to receive English-only reading instruction. To ensure a comprehensive understanding of this issue, first it is important to understand procedures used for identifying ELLs. In the state of Oregon, when enrolling their children in school, parents/guardians complete the Home Language Survey meant to establish the child’s primary language. All
children whose primary language is other than English are regarded as English language learners and are entitled to receive English-as-a-second-language (ESL) instruction. The Oregon Department of Education (ODE) requires that ELLs are placed for ESL instruction according to their level of English language proficiency. School districts select an assessment tool to determine the initial level of English language proficiency of ELLs.

The school uses the Woodcock-Muñoz Language Survey-English (LS-E) developed by Woodcock and Muñoz-Sandoval (1993b). This assessment tool will be described in more detail in Chapter 3. According to the Woodcock-Muñoz Language Survey-English (LS-E), individuals can demonstrate five levels of cognitive academic language proficiency (CALP): advanced English (5), fluent English (4), limited English (3), very limited English (2), and negligible English (1). Although they were identified as ELLs, children who demonstrated an advanced level of English language proficiency do not qualify to receive ESL services as their English proficiency is regarded as sufficient to function in the mainstream classroom on par with their English-speaking peers. In addition, parents/guardians of ELLs may exercise their right to opt out of ESL classes, although ODE still requires such children to participate in the English Language Proficiency Assessment (ELPA) until they demonstrate a sufficient level of English language proficiency. However, the combined number of students in these two categories, that is, ELLs with advanced English and ELLs not receiving ESL instruction on parental request, historically did not exceed 3-4 students accounting for approximately 5-6% of students identified as ELLs upon entry to kindergarten each year.

As the Woodcock-Muñoz Language Survey-English (LS-E) data were readily available, the school decided to use them to determine which ELLs should be placed to receive either L1 reading instruction or English-only reading instruction. This decision was based on an
assumption that ELLs with negligible (level 1) or very limited (level 2) English should be the ones to benefit from L1 reading instruction because their English language proficiency was not sufficient to receive English-only reading instruction. In addition, teacher observation and judgment during the first two months of kindergarten played a role in the placement. Children who did not appear to show sufficient progress in their English reading skills could be moved from the group receiving English-only reading instruction to the group that receiving L1 reading instruction. In summary, although ELLs were not placed either in L1 or English-only reading instruction on a random basis or through parental choice, this placement was not highly structured and, at times, even highly subjective. Moreover, no theoretical premises and/or research data were taken into account to support this approach to placement.

At the beginning of the 2007/2008 school year, the school decided to review this approach. The main goal of the review was to minimize the subjective aspect of the existing placement, make it more intentional, and, most importantly, to establish a theoretical rationale for the placement criteria. This theoretical rationale was provided by the Cummins (1979) threshold hypothesis. According to this hypothesis (it will be discussed in more detail in the Theoretical Framework section of this chapter), it is necessary to attain a certain level (threshold) of proficiency in a language in order to capitalize on the knowledge in this language. Therefore, if ELLs are placed in L1 reading instruction without regard to the level of L1 proficiency, their L1 proficiency may not be high enough to benefit sufficiently from learning to read in L1 first and then transferring this knowledge to L2 as their L2 proficiency increases. Consequently, the decision was made to consider the level of Spanish language proficiency as the primary criterion for placing ELLs either in L1 or English-only reading instruction.

For this purpose, in addition to screening ELLs for English language proficiency as
required by ODE, the school started to screen incoming kindergarten ELLs for Spanish language proficiency using the Woodcock-Muñoz Language Survey-Spanish (LS-S) developed by Woodcock and Muñoz-Sandoval (1993c). As a result, ELLs whose Spanish language proficiency is higher than their English language proficiency are placed in L1 reading instruction as well as those ELLs who scored limited (level 3) or fluent (level 4) in both English and Spanish. There may be a question as to why ELLs who scored at level 3 or level 4 in both English and Spanish are placed in L1 reading instruction versus English-only reading instruction. This category of ELLs is placed in L1 reading instruction based on the assumption derived from the threshold hypothesis. Cummins (1979) stated that “maintenance of L1 skills can lead to cognitive benefits for minority language children” (p. 232). ELLs who scored 3 or 4 in both languages may not only capitalize on their L1 knowledge to develop L1 reading skills for later transfer to L2 reading but, in addition, will get an opportunity to maintain their L1 through initial L1 reading instruction for a long-term positive impact on their cognitive development. From the perspective of the data analysis plan (it will be addressed in detail in Chapter 3), this change in the screening methodology can be viewed as an intervention that affected placement of kindergarten ELLs either in L1 or in English-only reading instruction. Therefore, given the fact that the ultimate goal of the bilingual program is to support English reading of ELLs, it could be informative for the researcher who has been teaching L1 reading to kindergarten for the past eight years and for the school administration to find out if focusing on the Spanish language proficiency when placing ELLs in two different types of reading instruction made a significant difference in terms of English reading achievement of all ELLs and particularly those ELLs who learned to read to L1. In addition, the use of the Oregon Assessment of Knowledge and Skills (OAKS) to measure English reading outcomes of third-grade students from 2007 to 2013
generated considerable data of test scores from six student cohorts: three cohorts before and three cohorts after the intervention. A flowchart of the bilingual program from 2004/2005 to 20012/2013 is provided in Table 1.

Table 1

*Flowchart of the Native Language Literacy (NLL) Program*

<table>
<thead>
<tr>
<th>Student Cohorts</th>
<th>School Years and Grades</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>04/05</td>
</tr>
<tr>
<td>Pre-Intervention 1</td>
<td>K</td>
</tr>
<tr>
<td>Pre-Intervention 2</td>
<td>K</td>
</tr>
<tr>
<td>Pre-Intervention 3</td>
<td>K</td>
</tr>
<tr>
<td><strong>Post-Intervention 1</strong></td>
<td><strong>K</strong></td>
</tr>
<tr>
<td><strong>Post-Intervention 2</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Post-Intervention 3</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Theoretical Framework**

*Common underlying proficiency (CUP) model.* The theoretical framework for this study was derived from the common underlying proficiency (CUP) model developed by Cummins (1980a) to support the concept of bilingual education. Apart from the level of linguistic proficiency and the language skills domains (speaking, listening, reading, and writing), another important element of research on bilingualism is the relationship between L1 and L2 of a bilingual person. There are two alternative concepts defining this relationship: separate underlying proficiency (SUP) and common underlying proficiency (CUP) (Cummins and Swain, 1986). The assumption that children coming to school lacking English language
(L2) proficiency need maximum exposure to the English language is aligned with the SUP model (Cummins & Swain, 1986). The SUP model implies that L1 and L2 are separate and unrelated, and there is a direct correlation between the amount of exposure to a language and the level of proficiency in that language (Cummins & Swain, 1986). Accordingly, if L1 and L2 proficiencies are separate and unrelated, knowledge and skills acquired in one language (either L1 or L2) cannot be transferred to the other language. Based on the SUP model, it appears illogical that any learning taking place in L1 can have an effect on later learning and/or knowledge in L2 (Cummins & Swain, 1986). The SUP model is illustrated in Figure 1.

![Figure 1](image-url)  

**Figure 1.** The separate underlying proficiency (SUP) model of bilingual proficiency. Adapted from “Bilingualism in Education,” by J. Cummins and M. Swain, 1986, p. 81. Copyright 1986 by Longman Group Limited.
Figure 2. The common underlying proficiency (CUP) model of bilingual proficiency. Adapted from “Bilingualism in Education,” by J. Cummins and M. Swain, 1986, p. 83. Copyright 1986 by Longman Group Limited.

Cummins (1980a) posited the common underlying proficiency (CUP) model. According to the CUP model, L1 and L2 proficiencies are seen as a common proficiency with two separate information entry channels for L1 and L2. The CUP model is illustrated in Figure 2. Cummins (Cummins & Swain, 1986) argued that the common underlying proficiency may not be immediately obvious and, at first sight, L1 and L2 proficiencies may appear to be independent and unrelated (as assumed in the SUP model). However, this separation exists only with respect to the surface manifestations of both languages (Cummins & Swain, 1986). These different surface manifestations of L1 and L2 can be compared to two tips of the same iceberg. While the two tips appear to belong to different icebergs, in reality they represent a “dual iceberg” where the major part of the iceberg is below the water level and is, therefore, invisible (Cummins & Swain, 1986). This submerged part of the iceberg is analogous to the common language proficiency. The “dual-iceberg” representation of bilingual proficiency is shown in Figure 3.
Figure 3. The “dual-iceberg” representation of bilingual proficiency. Adapted from “Bilingualism in Education,” by J. Cummins and M. Swain, 1986, p. 83. Copyright 1986 by Longman Group Limited.

**Linguistic interdependence.** The CUP model serves as a platform supporting other constructs developed by Cummins to enhance understanding of the theoretical rationale for bilingual education. One such construct is the interdependence principle which, according to Cummins (1979), is a central principle of bilingual education (Cummins & Swain, 1986).

Figures 2 and 3 provide an illustration of this principle: if L1 and L2 share a common underlying proficiency, it is logical to conclude that L1 proficiency and L2 proficiency are interdependent. The interdependence principle posits that to the extent that instruction in L1 is effective in promoting knowledge in L1, this knowledge in L1 will have an effect on knowledge in L2 (Cummins, 1981). Cummins (1981) specified two conditions that are critical for creating an education environment where the interdependence principle could be supported. The first condition is adequate exposure to L2 either at school or in the student’s environment in general. The second condition is adequate student motivation to learn L2.
Cummins’s interdependence principle contrasts with the “balance effect” principle (Macnamara, 1966). According to the “balance effect” principle, bilingual children advance their L2 skills at the expense of L1 skills, that is, the increase of L2 skills results in the decrease of L1 skills. Another construct that contrasts with the interdependence principle is the “linguistic mismatch” principle (Downing, 1974). This principle is based on the premise that bilingual children experience a mismatch between their home language (L1) and the language of school instruction (L2). Both the “balance effect” principle and the “linguistic mismatch” principle viewed bilingualism and, consequently, bilingual education as a source of “mental confusion” and “language handicaps” for learners (Cummins, 1979). Cummins & Swain (1986) asserted that, although the SUP model was not explicitly recognized by researchers who espoused these two principles, these principles appeared to be directly aligned with the SUP model. Evidence of this alignment is in the fact that if L1 and L2 proficiencies are separate and independent, as suggested by the SUP model, tapping into both proficiencies through bilingual instruction will lead to learners’ confusion. Therefore, it is possible to conclude that the CUP model in conjunction with the principle of linguistic interdependence asserting that “literacy-related aspects of bilingual proficiency in L1 and L2 are seen as common or interdependent across languages” (Cummins & Swain, 1986, p. 82) endorse the positive effect of the use of students’ L1 to support their learning in L2 in contrast to the SUP model that in conjunction with the “balance effect” and “linguistic mismatch” principles posit that the use of students’ L1 will have a negative effect on their learning in L2.

The interdependence principle is directly related to the construct of cross-linguistic or cross-language transfer (CLT). Cross-language transfer or cross-linguistic influence (CLI) has long been a founding concept with respect to learning a second language (Prieto Arranz, 2005).
According to Genesee, Geva, Dressler, and Kamil (2006), there have been two approaches to conceptualizing CLT. The first approach is represented by the interdependence principle by Cummins (1979). The second approach is represented by the contrastive analysis (Oldin, 1989). Genesee et al. (2006) noted that although the contrastive analysis acknowledges the possibility of CLT from L1 to L2, it is primarily concerned with the language structures, e.g., phonological, morphological, lexical, and syntactic, and, therefore, can be relevant to investigating CLT with respect to structural domains of literacy such as phonology (Genesee et al., 2006). Accordingly, as argued by Genesee et al. (2006), the contrastive analysis fails to account for CLT with respect to more metacognitive literacy constructs such as reading comprehension. In difference to the contrastive analysis, the construct of linguistic interdependence that was derived by Cummins from the CUP model conceptualizes the cross-linguistic transfer of skills that are more cognitive in nature. For that reason, the principle of linguistic interdependence can provide a better understanding of how L1 literacy can support L2 literacy.

**Cognitive academic language proficiency.** The interdependence principle articulated by Cummins (1981) in conjunction with the construct of CLT postulated that L1 and L2 are developmentally interdependent, that is, L1 development can facilitate L2 development (Genesee et al., 2006). At the same time, Cummins (2000) cautioned that not all aspects of L1 are equally instrumental in L2 development. This statement is supported by two other constructs developed by Cummins (1980b): basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP). Cummins (1980b) stated that “there has been relatively little inquiry into what forms of language proficiency are related to the development of literacy proficiency in school contexts, and how the development of literacy proficiency in L1 relates to the development of literacy proficiency in L2” (p. 27). Cummins (1984) pointed out that
although educators and psychologists generally expected children appearing to be fluent in the English language to demonstrate success on English language academic tasks, such children frequently performed poorly. The distinction between the constructs of BICS and CALP helps to understand this phenomenon. Although such children already achieved an adequate level of basic interpersonal communicative skills (BICS) in L2, they can still be lacking in cognitive academic language proficiency (CALP) in L2.

Prior to the articulation of the BICS/CALP distinction by Cummins (1980b), Oller (1979) attributed differences in everyone’s language proficiency to the construct of global language proficiency indicating that it was responsible for “the reliable variance in a wide variety of language proficiency measures (p. 413). According to Cummins (2000), Oller conceptualized the language proficiency as “a unitary global dimension that was largely indivisible from intelligence” (p.119). In contrast to Oller (1979), Cummins (1979) argued that not all differences in language use and performance could be attributed to the existence of the global language proficiency. Cummins (1980b) pointed out that all children, with the exception of those who are severely retarded or autistic, are able to acquire basic interpersonal communicative skills. This assumption correlated with Chomsky’s (1965) position that basic phonological, syntactic, and lexical skills needed for interpersonal communicative situations are characteristic of all native speakers. Therefore, instead of one general construct of global language proficiency Cummins (1979) distinguished two separate constructs: basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP). Cummins (1979) suggested that the term “cognitive academic language proficiency” should be used in lieu of “global language proficiency” to make a distinction between the language proficiency related to literacy skills and the language proficiency related to cognitively undemanding interpersonal situations.
Cummins (1980b) demonstrated the difference between BICS and CALP by adapting Shuy’s (1978) “iceberg metaphor” shown in Figure 4. The “visible” part of the iceberg represents more obvious language proficiencies (pronunciation, grammar, and vocabulary) that are manifested in the context of everyday communicative language use. The “hidden” part of the iceberg represents the type of language proficiency required for performing cognitive academic tasks. Cummins (1980b) maintained that a distinction similar to that between BICS and CALP had been made by other theorists. For example, Burt and Dulay (1978) distinguished between “natural communication” tasks and “linguistic manipulation” tasks. However, by “linguistic manipulation” Burt and Dulay (1978) primarily meant metalinguistic awareness, while Cummins regarded metalinguistic awareness as only one of the many aspects of CALP. Cummins (1980b) further advanced the constructs of BICS and CALP by demonstrating their developmental trends.
as shown in Figure 5. As can be seen from Figure 5, CALP follows the general trajectory of cognitive development that starts to plateau in mid-adolescence, while the formation of BICS is usually completed by the time children reach five or six years of age. According to Cummins (1980b), another important difference between the two types of language proficiency is that while CALP is strongly related to academic success, BICS is largely not related to it. Most important, in the view of Cummins (1980b), is the fact that both with respect to L1 and L2, the level of BICS cannot provide any indication about the level of CALP. Therefore, ELLs that have a native-like proficiency in L2 BICS can struggle in academic areas that require L2 CALP.

![Figure 5](image_url)  

Using the constructs of BICS and CALP Cummins (1980b) extended the interdependence principle specifically to cognitive/academic aspects of L1 and L2. In doing so, Cummins (1979) posited that L1 and L2 CALP are interdependent and that L2 CALP is partially a function of L1 CALP. Cummins (1980b) contended that earlier learning of literacy skills in L1 can predict later learning of these skills in L2. Figure 6 illustrates the assertion that while L1 and L2 can be
significantly different in terms of “surface” features (e.g., phonology, syntax, and lexicon), L1 and L2 share a common cognitive/academic proficiency that determines performance on cognitive and academic tasks, such as reading, in both languages (Cummins, 1980b).

![Diagram of bilingual proficiency]

*Figure 6.* The “dual-iceberg” representation of bilingual proficiency. Adapted from “The entry and exit fallacy in bilingual education” by J. Cummins, 1980, NABE Journal, 4(3), p. 36. Copyright by Routledge.

Cummins (1980b) cautioned that the linguistic interdependence and cross-linguistic transfer were conditional on a number of factors that could influence the relationship between L1 and L2 cognitive academic proficiencies. These factors include motivation to maintain L1 and/or learn L2 as well as the level of exposure to L1 and L2. The examination of factors that can influence the degree to which academic knowledge developed in L1 supports the development of academic knowledge in L2 led to the formulation of the threshold hypothesis.

**Threshold hypothesis.** The threshold hypothesis suggested that there may be threshold levels of linguistic competence that bilingual children need to attain in both languages so that potentially beneficial aspects of bilingualism, such as linguistic
interdependence and cross-linguistic transfer, could come into effect (Cummins & Swain, 1986). Accordingly, levels of linguistic competence may determine the extent to which a bilingual learning experience may influence cognitive development in bilingual children (Cummins & Swain, 1986). Cummins (1979) argued that, according to the threshold hypothesis, the levels of competence in L1 and L2 are able to mediate the interaction of a bilingual child with the language environment with respect to both input and output. To gain a better understanding of how levels of bilingual competence may affect cognition, it is necessary to distinguish two levels of threshold (Cummins, 1979). A lower threshold level implies that a sufficiently weak level of competence in a language can negatively affect an interaction between the child and a learning environment in this language. Cummins (1979) maintained that the lower threshold level was not absolute and depended upon the stage of cognitive development and academic demands. The achievement of the lower threshold helps to avoid the negative effects of a low level linguistic competence on cognitive functioning. Cummins (1979) emphasized that the low levels of competence bilingual children often exhibit in both languages due to not yet having achieved the lower threshold validated the existence of this threshold level.

In addition, Cummins (1979) observed that a native-like level in at least one of the languages was the prerequisite for the higher threshold level of bilingual competence. Therefore, attainment of this higher threshold level may depend on the maintenance of L1 skills. While Toukomaa and Skutnabb-Kangas (1977) claimed that the higher threshold level did not have either negative or positive effect, Cummins (1979) disputed this view stating that research supported the premise that “the maintenance of L1 skills can lead to cognitive benefits for minority language children” (p. 232). Consequently, the threshold hypothesis advances
understanding of the relationship between L1 and L2 by explaining how different levels of bilingualism may influence academic development in bilingual children and notably positing that capitalizing on the higher threshold level in L1 can be beneficial for ELLs.

**Summary.** The seminal works by Cummins on the threshold hypothesis and the CUP model together with the constructs of linguistic interdependence and CALP that both extend the CUP model created a comprehensive theoretical framework supporting various forms of bilingual education, including L1 reading instruction. The theoretical premises outlined by Cummins contribute to the understanding of the relationship between L1 and L2 and the role L1 can specifically play in the academic development of ELLs in areas such as English reading. Accordingly, the CUP model and the threshold hypothesis were the logical choice for the theoretical rationale of this study set in the context of a bilingual program and aimed at exploring factors influencing the English reading achievement of elementary ELLs.

**Research Questions**

The study posed two research questions and tested the hypotheses derived from the research questions:

**RQ1:** How, if at all, did the implementation of a new kindergarten screening procedure for Spanish-speaking ELLs regarding placement in L1 versus English-only reading instruction during the first four years of elementary school affect third-grade student performance on the OAKS standardized reading achievement test?

**H1:** Following a change in the screening procedure for Spanish-speaking ELLs, third-grade reading achievement will significantly improve relative to that of students prior to the implementation of the new screening procedure.

**RQ2:** How, if at all, did the implementation of a new kindergarten screening procedure
for Spanish-speaking ELLs affect third-grade student performance on the OAKS standardized reading achievement test for those placed in L1 versus English-only reading instruction?

H2: Following a change in the kindergarten screening procedure for Spanish-speaking ELLs, third-grade reading achievement test scores of students who received L1 reading instruction will be significantly higher than those of students who received English-only reading instruction.

**Significance of the Study**

The significance of this research study is underscored by the fact that ELLs constitute the fastest growing segment of student population in the United States (Calderón, Slavin, & Sanchez, 2011). Based on the data from the National Clearinghouse for English Language Acquisition (NCELA, 2002) Cárdenas-Hagan, Carlson, and Pollard-Durodola (2007) concluded that between 1990 and 2004 school enrollment of ELLs doubled with Spanish-speaking ELLs accounting for about three fourth of this growth. At the same time, many ELLs continue to be at risk for reading failure and school drop-out (Gottardo, 2002). Citing the data from the National Center for Education Statistics (2005) Cárdenas-Hagan et al. (2007) noted that “66% of the ELL population in the United States scored below the basic reading level in fourth grade and 67% in eighth grade” (p. 249). These data demonstrate that both researchers and practitioners need to be actively exploring the issue of ELL English reading underachievement and the ways to close this gap. As a first step to closing this gap research should focus on bilingual reading programs that, according to Goldenberg (2008), enroll about 50% of ELLs in the United States.

In summary, there are three factors that determine the practical significance of the examination of bilingual programs. The first factor is a large proportion of ELLs in schools
throughout the United States. The second factor is documented underachievement of this category of students in English language reading in conjunction with potential negative consequences of this underachievement such as higher school dropout rates. The third factor is the fairly wide use of bilingual programs despite recent federal and state policies restricting the amount of instructional time in students’ L1 (Slavin et al., 2011). The fact that 80% of ELLs are native Spanish speakers (Capps, Fix, Murray, Ost, & Passel, 2005) makes the examination of Spanish language bilingual programs particularly relevant to the understanding of how to ensure reading success of ELLs in general and of Spanish speaking ELLs in particular. This consideration is enhanced by the fact that Spanish-speaking ELLs are the fastest growing ethnic minority group in the United States as they account for three fourths of the overall increase in the enrollment of ELLs (Cárdenas-Hagan et al., 2007) and that bilingual programs are predominantly based on the Spanish language (Slavin & Cheung, 2005).

Baker et al. (2012) pointed out that there is clear need for more research related to L1 instruction for ELLs. This need is underscored by the following factors: (1) the number of existing studies is insufficient to support the premise that L1 reading instruction has beneficial outcomes for ELLs; (2) many of the studies are dated and do not reflect the current situation in education; and, (3) existing studies do not always provide sufficient information about the organization and delivery of L1 reading instruction (e.g., Baker, S. K., & Baker, D. L., 2008; Francis et al., 2006; Gersten & Baker, S. K., 2000a, 2000b). Consequently, these factors further support the significance of this study. Finally, the theoretical significance of the study lies in the fact that “the validity of any theoretical principle is assessed precisely by how well it can account for the research findings in a variety of contexts” (Cummins, 1999, p. 26). Cummins (1999) argued that considerably more research was needed to refine theoretical principles of bilingual
education “to take account of the multiple interactions that occur among variables that contribute to bilingual students’ academic success (p. 32).

**Definition of Terms**

In this section, key terms that are central to the understanding of the context and content of the present research study are conceptually and operationally defined. Sources are cited as appropriate.

Basic interpersonal communicative skills (BICS) are the skills needed in cognitively undemanding interpersonal situations (Cummins, 1979).

Bilingual education differs fundamentally from English immersion. In English immersion, ELLs “are expected to learn in English from the beginning, and their native language plays little or no role” (Slavin & Cheung, 2005, p. 250). Bilingual education gives ELLs “significant amounts of instruction in reading and/or other subject in their native language” (Slavin & Cheung, 2005, p. 250).

Bilingualism is defined as linguistic proficiency in two languages. Different authors set different standards of language proficiency required to qualify as a bilingual: from at least a minimal level of language ability (Macnamara, 1967) to complete mastery of two different languages (Oeistreicher, 1974), to native-like proficiency in two or more languages (Bloomfield, 1933).

Common underlying proficiency (CUP) model posits that L1 and L2 share a common proficiency, but each language has a separate information entry channel (Cummins & Swain, 1986).

Cognitive academic language proficiency (CALP) is the language proficiency required to perform cognitively demanding academic tasks (Cummins, 1979).
Cross-linguistic transfer (CLT), or cross-language transfer, is defined as “the influence of the mother tongue (L1) on the learner’s performance in and/or development of a given target language” (Sharwood Smith, 2014, p. 13).

English language learners (ELLs) are also referred to as second language learners or minority language children (e.g., Cummins, 1979). English Language Learners (ELLs) are defined as “students who do not understand enough English to learn without support in mainstream classrooms” (Russakoff, 2011, p. 4).

First language (L1) is also referred to as native language (e.g., Slavin & Cheung, 2005), home language, or primary language.

Interrupted times series (ITS) analysis is “series of observations made on the same variable consecutively over time” (Shadish, Cook, & Campbell, 2002, p. 172). Interrupted time series compare observations before and after an identifiable event with the purpose to examine the impact of this intervention (Tabachnick & Fidell, 2007).

Language proficiency refers to a person’s ability to use a language for a variety of purposes, e.g., speaking, listening, reading, and writing (American Council on the Teaching of Foreign Languages, 2012).

Native Language Literacy (NLL) is a bilingual reading program for Spanish-speaking ELLs in grades K through 3 at the elementary school that was the research site for the study.

Oregon Assessment of Knowledge and Skills (OAKS) is an element of the Oregon Statewide Assessment System (OSAS). It includes reading, mathematics, science, and social studies tests administered in grades 3 to 8 and grade 10.

Oregon Department of Education (ODE) developed and administered annual OAKS standardized achievement tests mandatory for state public schools.
Second language (L2), in the context of this study, is also referred to as the English language.

Separate underlying proficiency (SUP) model posits that L1 and L2 are not connected and, therefore, are unrelated (Cummins & Swain, 1986).

Transitional bilingual education (TBE) is a specific type of bilingual programs where children are taught to read primarily or entirely in their first L1 and then transition to English reading instruction. In early-exit models, transition to English is completed in second or third grade. In late-exit models, students continue to receive L1 instruction throughout elementary school (Slavin et al., 2011).
CHAPTER 2: REVIEW OF THE LITERATURE

Introduction

The goal of this study was to explore the effect of a change in the screening methodology used for placing students in L1 or English-only reading instruction on the English reading achievement of elementary ELLs and thereby to contribute to the research literature on the bilingual education of ELLs. The literature review relates the existing research to the variables in this research study. This review reveals that no research examining the effect of changes in the screening criteria used for placing ELLs in bilingual instruction (independent variable) as opposed to English immersion has been published to date. However, there is an abundance of research with respect to L2 literacy in ELLs at the primary grade levels, that is., the language of literacy instruction for ELLs and the effects of L1 literacy instruction on L2 literacy (e.g., Branum-Martin et al., 2010; Manis, Lindsey, & Bailey, 2004; Mathes, Pollard-Durodola, Cárdenas-Hagan, Linan-Thompson, & Vaughn, 2007). This body of research can be related to the L2 reading achievement of ELLs which is the dependent variable in this investigation. Accordingly, by focusing on the foregoing types of research studies, this literature review provides a context for the design of the study and underscores the fact that additional research is needed “to more fully explain how various interventions affect the development of English reading skills among English language learners” (Slavin & Cheung, 2005, p. 275).

The review of literature examines the existing research from four different perspectives. First, it looks into the issue of English literacy development in ELLs. Second, it examines effective methods of literacy instruction of ELLs. Third, it reviews research on the language of reading instruction for ELLs. Finally, it analyzes the research data on cross-linguistic transfer of reading skills from L1 to L2. All of these perspectives are considered in terms of the role
reading instruction in L1 may play in the development of English literacy. The literature review concludes with a thesis statement about the role of L1 reading instruction with respect to English literacy development in ELLs and with a summation outlining the implications of the thesis statement for research and practice.

**Second Language Literacy Development in ELLs**

**Literacy development in ELLs and native speakers.** There is a growing awareness of the fact that English literacy development in ELLs should not be equated to English literacy development in native speakers (e.g., Bernhardt, 2003; Escamilla, 2009; Hammer, Jia, & Uchikoshi, 2011). In 2006, the National Literacy Panel on Language-Minority Children and Youth (NRP) produced a report on literacy development in second language learners (August & Shanahan, 2006). This report was written in response to a growing concern about the literacy underperformance in this group of students and is based on an examination of the research on the acquisition of second language literacy conducted by the panel. In Escamilla’s 2009 review of the report, the author discussed how the development of reading skills in ELLs is different from and similar to that of native speakers.

According to Escamilla (2009), the panel reported parallels in the research findings from the studies of literacy development in ELLs and monolingual students. One similarity pointed out by Escamilla (2009) is that, with respect to word-level reading skills, e.g., decoding, it is quite possible for ELLs to achieve the same level of proficiency as that of their English-speaking peers. However, Escamilla (2009) noted that in terms of text-level reading skills, e.g., reading comprehension, the panel reported that these skills in ELLs can seldom be comparable to those of native speakers. Escamilla (2009) argued that because the panel’s findings demonstrated similar proportions of struggling readers among both ELLs and monolingual students, one
should not conclude that being an ELLs would necessarily lead to being a struggling reader.

**Monolingual approach to second language literacy.** One concern expressed by Escamilla (2009) was related to the monolingual approach taken by the panel. Escamilla (2009) observed that the panel viewed monolingual instruction as a norm and, therefore, described ELLs in “deficit terms” as those who need to “catch up.” Escamilla (2009) reasoned that now when bilingualism is the norm it should have been addressed in the report. Escamilla (2009) concluded that, by adopting the monolingual approach, the panel failed to take into account the strengths and abilities of ELLs.

The conclusion by Escamilla (2009) was endorsed by Hammer et al. (2011) who claimed there is only a limited understanding of how ELLs acquire their literacy skills and that studies including ELLs seldom take into account their native language abilities. Hammer et al. (2011) examined what is currently known about the literacy development in two major sub-groups of second language learners: Latinos and Asians. By using the term “dual language learner” Hammer et al. (2011) underscored the fact that L1 skills are relevant to the development of L2 literacy. The authors pointed out that the growth in receptive Spanish language skills in preschool children was a positive predictor of English literacy development in kindergarten and first grade. According to Hammer et al. (2011), these data support the premise of linguistic interdependence posited by Cummins (1979) which states that children’s knowledge in one language helps to develop their knowledge in another language.

**Need for additional research on second language literacy.** The report on developing literacy in ELLs synthesized the data from the empirical research published between 1980 and 2002 (Escamilla, 2009). However, the concept of L2 literacy has been addressed in reading research since the second half of the nineteenth century (Bernhardt, 2003). Having noted the
interest of researchers in L2 reading, Bernhardt (2003) pointed out that historically L1 and L2 literacies have been regarded as essentially the “same.”

Consequently, according to Bernhardt (2003), research has mostly focused on English language reading. Bernhardt (2003) disputed this approach, claiming that “the mere existence of a first language …renders the second-language reading process considerably different from the first-language reading process because of the nature of information stored in memory” (p. 112). Bernhardt’s (2003) argument was based on an assumption that with respect to L2 readers the reading process is influenced by memory-based representations from their L1. According to Bernhardt (2003), these “representations in memory possess varying degrees of usefulness and relatedness for cognitive processing” (p. 113).

**Conclusion.** The evidence presented by Bernhardt (2003), Escamilla (2009), and Hammer et al. (2011) suggests that learning to read in L2 should not be viewed and studied solely from the monolingual perspective. This view underscores the point that to close the reading achievement gap between ELLs and native speakers it is important to look beyond instructional practices that have demonstrated effectiveness with respect to the reading achievement of monolingual students.

**Effective Reading Instruction for ELLs**

**Need for a specific approach to effective reading instruction for ELLs.** There is a considerable body of research on the issue of what should be regarded as effective English reading instruction for ELLs (e.g., Bernhard, 2003; Calderón et al., 2011; Fitzgerald, 2000; Lenters, 2004; Manyak, 2007; Mathes et al., 2007). The assumption that L1 and L2 literacy development processes should not be viewed as identical, prompted Bernhardt (2003) to consider the need for unique approaches to L2 reading instruction. Bernhardt (2003) disagreed with the
position taken by Fitzgerald (2000) that there are insufficient research findings to justify the need for developing a special theory of L2 reading. Bernhardt (2003) argued that the view adopted by Fitzgerald (2000) would encourage both teachers and researchers to ignore language and literacy differences demonstrated by ELLs.

Bernhardt’s (2003) argument may explain why some authors, such as Mathes et al. (2007) and Lenters (2004), have also noted that there is insufficient research upon which to base instructional decisions with respect to effectively addressing the needs of ELLs in reading development. Mathes et al. (2007) observed that in the 2000 report on the assessment of the research literature on reading prepared by the National Reading Panel it was explicitly stated that assessment goals did not involve focusing on any specific population such as ELLs.

**Elements of effective reading instruction for ELLs.** Other researchers (e.g., Manyak, 2007; Calderón et al., 2011) have argued that it is possible to identify specific elements of effective English reading instruction for ELLs. Calderón et al. (2011) promoted certain salient instructional strategies effective for ELLs in all grades. These strategies included integration of language, literacy, and content, cooperative learning, tutoring, and other interventions for struggling readers. Calderón et al. (2011) claimed that these strategies are critical regardless of the language of reading instruction.

Manyak (2007) suggested that reading success in all ELLs could be achieved with the use of a “multifaceted form of classroom instruction” (p. 197). Manyak (2007) outlined a four-pronged approach that includes explicit code and comprehension instruction, language-rich instruction, socio-culturally informed instruction, and additive literacy instruction. Manyak (2007) stressed the importance of “additive literacy instruction” in light of existing research confirming that the development of literacy skills either in L1 or L2 has a positive effect on
literacy achievement in the other language. The logical conclusion derived from the argument made by Manyak (2007) is that, for ELLs, the language of reading instruction may be an essential factor to address when considering approaches to facilitate and enhance reading achievement in this population of learners.

Indeed, having noted insufficient research to guide decision-making in the area of supporting struggling readers who are also ELLs, Mathes et al. (2007) emphasized that most of the existing research focused on the language of reading instruction for ELLs (English vs. native language) and the timing of the transition to English-only reading instruction (early vs. late). According to Mathes et al. (2007), while the language of reading instruction for ELLs is a justifiably important component of the discussion about ensuring reading success in ELLs, the resolution of the language-of-instruction issue “will do little to inform us about how to promote reading success among ELLs who struggle to learn to read, regardless of the language of instruction” (p. 261).

Mathes et al. (2007) examined the results of four large-scale studies on preventing reading failure in Spanish-speaking ELLs. The studies have been conducted with the purpose of establishing a set of conditions that would lead to effective development of reading and writing skills in Spanish-speaking ELLs. Mathes et al. (2007) concluded that it was important to teach foundational reading skills in conjunction with complex cognitive skills. Mathes et al. (2007) determined that it is a challenge “to integrate foundational and complex cognitive skills with student’s language and culture” (p. 269). At the same time, the researchers noted that such integration could have a positive effect on reading outcomes in Spanish-speaking ELLs who are struggling with English reading.
**Conclusion.** Despite a considerable number of studies on the effective reading instruction for ELLs, some authors (e.g., Mathes et al., 2007; Lenters, 2004) have insisted there is still not enough empirical data to provide clear guidelines on instructional practices to alleviate reading difficulties experienced by ELLs. Other authors (e.g., Hedgcock & Atkinson, 1993; Fitzgerald & Noblit, 1999, 2000; Fitzgerald, 2000) have disputed the need for specialized approaches to the reading instruction of ELLs.

A middle-of-the-road position was taken by such researchers as Manyak (2007) and Calderón et al. (2011). These researchers advocated a multi-faceted approach to the reading instruction for ELLs. Accordingly, they outlined their vision of effective reading instruction for ELLs. However, these authors took different positions with respect to the language of instruction. While Manyak (2007) acknowledged the relevance of the language-of-instruction issue, Calderón et al. (2011) maintained that effective instructional strategies were not necessarily related to the language of instruction.

**Language of Reading Instruction for ELLs**

**The role of L1 as the language of reading instruction.** The debate about the language of literacy instruction for ELLs has been ongoing and “fierce” for several decades (Calderón et al., 2011). Some studies (e.g., Mathes et al., 2007; Calderón et al., 2011) emphasized the fact that the language of reading instruction for ELLs should not take precedence in the discussion about effective instructional techniques to assist ELLs in becoming proficient English readers. Other researchers (e.g., Slavin & Cheung, 2005) disagreed with this view stating that the role of the native language in the reading instruction of ELLs is clearly the dominant issue. At the same time, Slavin and Cheung (2005) maintained that literacy success in ELLs is determined by multiple aspects of instruction. The view of Slavin and Cheung (2005) on the role of the native
language in English reading development was endorsed by Domínguez de Ramírez and Shapiro (2007) whose research established that initial L1 reading instruction supported later L2 literacy.

Slavin and Cheung (2005) noted that throughout the 1970s and 1980s there was a strong drive for bilingual education. Consequently, in elementary schools, second language learning children were often taught partially or entirely in their L1 before being mainstreamed to English-only classrooms. Slavin and Cheung (2005) went on to say that although bilingual programs are still fairly common, current federal policies set a limit as to the amount of instruction that can be delivered in students’ L1. Additionally, some states have scaled down bilingual programs due to new state policies.

An ongoing discussion about the role language of instruction plays in the English literacy development of ELLs has not resulted in a clear consensus in the research community. While some researchers (e.g., Domínguez de Ramírez & Shapiro, 2007; Slavin & Cheung, 2005) argued the importance of the issue and asserted a positive effect of L1 reading instruction on English reading, others (e.g., Mathes et al., 2007; Calderón et al., 2011) did not regard the role of the language of instruction as highly significant or critical. The latter approach appears to be reflected in the federal and state policies that over recent decades limited the amount of instructional time ELLs could receive in their primary language (Slavin & Cheung, 2005; Calderón et al., 2011).

**Potential drawbacks and benefits of L1 reading instruction.** In their 2005 meta-analysis of research on the language of reading instruction for ELLs Slavin and Cheung pointed out that reviews of educational outcomes of L1 reading instruction arrived at conflicting conclusions. According to Slavin and Cheung (2005), the central argument used by proponents of bilingual instruction is that, to begin with, children should receive reading instruction in L1
until they are sufficiently proficient in English which should usually occur in third or fourth grade. The strategy would help ELLs to avoid the “failure experience” when they are required to simultaneously develop oral English language proficiency and learn to read in English.

Slavin and Cheung (2005) reported multiple studies that support the premise that the level of reading proficiency in L1 is a strong predictor of children’s reading performance in the English language. Based on the results of their longitudinal study Lindsey, Manis, and Bailey (2003) also concluded that there is strong “cross-linguistic prediction of reading skills in bilingual children” (p. 493). Similarly, the study by Flood, Lapp, Villamil Tinajero, and Rollins Hurley (1997) cited empirical evidence to the effect that ELLs who became proficient in L1 literacy developed academic and literacy skills more rapidly in both English and L1.

On the other hand, Slavin et al. (2011) noted that the main argument used by opponents of bilingual education was that L1 instruction caused delays in English language development. Rossell and Baker (1996), for example, maintained that greater emphasis on the English language in English-only reading programs for ELLs should translate into higher achievement in English reading. While studying the language of literacy instruction on Spanish-English bilinguals, Rolla San Francisco, Mo, Carlo, August, and Snow (1996) argued that “formal literacy instruction in English proved to be the most powerful overall” (p. 639). This conclusion correlates with the view by Lesaux and Siegel (2003) who established that ELLs in a high-quality early English reading program were able to match and even exceed the reading development of their monolingual peers.

**Bilingual and immersion reading programs.** Slavin and Cheung (2005) noted that English immersion and bilingual programs were the two most common types of programs used for addressing the problem of how young ELLs starting kindergarten or first grade can increase
their English language proficiency while mastering required academic content. In English immersion programs, as described by Slavin and Cheung (2005), from the very beginning students receive all instruction in English, and their primary language, beyond occasional translation or explanation, hardly plays any role in their education. In contrast to English immersion, bilingual programs provide instruction in reading (and/or other subject areas) in the primary language. According to Slavin and Cheung (2005), most bilingual programs in the United States use Spanish language. In addition to Slavin and Cheung (2005), the terms “immersion” and “bilingual” programs have also been used with the same definition by other authors, such as Flood et al. (1997), Branum-Martin et al. (2010), and Rolstad et al. (2005).

Arguing the importance of bilingual education for overall academic development of ELLs, Flood et al. (1997) contended that American public schools have historically been amenable to incorporating languages other than English into the instructional process. Noting that the debate about the merits of immersion programs versus bilingual programs has been going on for several decades, Flood et al. (1997) pointed out that there was “ample evidence” in support of bilingual education and literacy in L1 for academic success of ELLs.

In a meta-analysis of 17 experimental studies comparing bilingual and English-only reading programs, Slavin and Cheung (2005) reported a lack of high-quality research comparing the effects of such programs. The review by Slavin and Cheung (2005) determined that 12 studies found effects in favor of bilingual programs while five studies did not observe any differences between the effects of bilingual and English immersion programs and “none of the studies found results favoring English immersion” (p. 273). This statement is in line with an earlier one by Flood et al. (1997) about the “ample evidence” in support of bilingual programs. In a 2010 report on the evaluation of bilingual programs, Branum-Martin et al. also noted that
“native language programs generally have positive impacts for English literacy” (p. 351). Rolstad et al. (2005) arrived at a similar conclusion in their meta-analysis of the effectiveness of bilingual and English immersion programs.

A multi-year randomized evaluation of transitional bilingual education (TBE) by Slavin et al. (2011) defined this education as programs where students are first taught to read in their primary language and later transition to English reading instruction. Slavin et al. (2011) called this period of time when ELLs receive literacy instruction in the native language a “hallmark” of TBE. The term “transitional” also reflects the fact that, in TBE, initial reading instruction in L1 is combined with teaching oral English language and other subjects in English from the beginning of children’s school experience. Slavin et al. (2011) commented that, according to proponents of TBE, beginning L1 literacy instruction was expected to help Spanish-speaking children become better readers in English. The data from the study by Slavin et al. (2011) did not validate this expectation.

Nonetheless, a most recent study by Lindholm-Leary (2014) aimed at better understanding of bilingual and bi-literacy skills in young Spanish-speaking children from families with low socioeconomic status arrived at the conclusion that bilingual instruction of such children in preschool and early elementary grades was advantageous for this category of students. Slavin et al. (2011) pointed out that one of main arguments of the opponents of bilingual education was that bilingual instruction interferes with or even delays English language development in ELLs. Contrary to this position, the findings of the study by Lindholm-Leary (2014) confirmed that “English language development was not negatively impacted by bilingual instruction” (p. 157). Therefore, Lindholm-Leary (2014) asserted that “there was no significant disadvantage to being instructed bilingually and no significant advantage to being instructed
Conclusion. Existing literature demonstrates a variety of conflicting opinions on the drawbacks and benefits of L1 instruction. On the one hand, the literature has shown definite benefits of L1 instruction, such as avoiding initial failure experiences and helping to accelerate students’ English reading proficiency (e.g., Slavin & Cheung, 2005; Flood et al., 1997). An additional benefit of initially learning to read in L1 is that the level of proficiency in this area can predict later achievement in English reading (e.g., Slavin & Cheung, 2005; Lindsey et al., 2003). On the other hand, not all researchers acknowledge these benefits, claiming that L1 reading instruction may be detrimental to English reading development or redundant because an emphasis on English reading instruction can be sufficient to produce English reading success in ELLs (e.g., Rolla San Francisco et al., 2006; Lesaux & Siegel, 2003).

Research appears to be fairly consistent in confirming that bilingual education programs generally provide more benefits to ELLs than English immersion programs (e.g., Branum-Martin et al., 2010; Flood et al., 1997; Rolstad et al., 2005; Slavin & Cheung, 2005. However, in contrast to that statement, the study on transitional bilingual programs as a specific type of bilingual programs by Slavin et al. (2011) did not demonstrate strong positive results with respect to English reading in ELLs. This result is partially consistent with the data from an earlier study by Slavin and Cheung (2005) suggesting that in some cases research did not detect any difference between the effects of bilingual and immersion programs.

Cross-Linguistic Transfer in Literacy Development

Theory of cross-linguistic transfer. Cross-linguistic relationships in ELLs were one of the issues addressed in the 2006 report prepared by National Literacy Panel on Language-Minority Children and Youth. Escamilla (2009) noted that the studies reviewed in the report
were based on the theory of transfer. The theory of transfer is related to the linguistic interdependence principle developed by Cummins (1979). Cummins (1979) hypothesized that language and literacy skills are transferrable from L1 to L2. Various authors have used the term “cross-linguistic transfer” (e.g., Atwill, Blanchard, Gorin, & Burstien, 2007; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2008; Proctor, Mo, Carlo, August, & Snow, 2006) or “cross-language transfer” (e.g., Manis et al., 2004; Durgunoğlu et al., 1993) to refer to the transfer of various literacy skills from L1 to L2.

Evidence of cross-linguistic transfer. Atwill et al. (2007) stated that one of the fundamental meta-linguistic skills critical for early reading development is phonemic awareness. Atwill et al. (2007) found evidence that there is cross-linguistic transfer (CLT) of phonemic awareness in kindergarten students with average and above-average Spanish language proficiency. No significant CLT was observed in children with language skills below average. Atwill et al. (2007) attributed this difference to the fact that CLT may be dependent on a certain level of language competency in L1. According to Atwill et al. (2007), the implication of these findings in terms of the benefits of initial L1 literacy instruction is that Spanish-speaking children with low levels of L1 ability may not benefit from the CLT of L1 phonemic awareness to L2 literacy. This observation suggests that “a one-size-fits-all approach” to bilingual programs should not be expected to produce competent readers in L2, a conclusion that echoes the argument by Flood et al. (1997) that English immersion would not meet the needs of all ELLs in reading instruction.

The results obtained by Atwill et al. (2007) supported earlier findings by Durgunoğlu et al. (1993) who observed CLT of phonological awareness in young children as well as CLT of other literacy skills. Durgunoğlu et al. (1993) further suggested that evidence of CLT or the lack thereof could be used as a diagnostic tool for predicting cognitive and learning difficulties.
Children who received reasonable exposure and instruction in L1, but failed to demonstrate CLT to L2 could be experiencing cognitive and/or learning challenges that would negatively affect literacy development in both languages.

Proctor et al. (2006) examined CLT in relation to L2 reading comprehension. The results of the study supported the premise that there is transfer of reading skills from L1 to L2. However, the study detected limits to such transfer. For example, there was limited evidence for the influence of Spanish vocabulary knowledge on English comprehension. Therefore, the researchers concluded that while the development of reading skills in L1 in conjunction with literacy instruction focused on the development of L2 is beneficial, L1 development does not automatically translate into better English reading comprehension. These findings are in contrast to the findings by Atwill et al. (2007) who established a clear link between the level of early L1 skills and CLT of phonemic awareness.

Other studies examined CLT with respect to different components of literacy and demonstrated an array of results as to the extent of CLT. For example, in the study by Mathes et al. (2007) undertaken for the purpose of providing information on effective reading instruction for Spanish-speaking elementary students, the researchers reported no evidence of cross-linguistic transfer. Further, the study by Manis et al. (2004) examining the development of reading skills in both languages from kindergarten to second grade in Spanish-speaking ELLs noted only “a modest amount of cross-language transfer” (p. 214).

The study by Gabriele, Troseth, Martohardjono, and Otheguy (2009) examined cross-linguistic relationships from yet another perspective. Their study found that L1 syntactic development strongly influenced text-level listening comprehension in L2. These findings paralleled the findings of Lindsey et al. (2003) who determined that L1 phonological awareness
both influenced and predicted word identification skills in L2. The longitudinal study of Spanish and English reading by Lindsey et al. (2003) demonstrated the presence of CLT in predictor and outcome variables with respect to such literacy skills as letter and word knowledge, print awareness, and comprehension. The researchers claimed that their findings broadened the range of literacy variables where CLT was observed.

Sparks et al. (2008) used a variety of measures, such as phonological awareness, word decoding, and reading comprehension, to examine CLT from the perspective of early L1 predictors of later L2 reading skills. A significant conclusion by Sparks et al. (2008) was that L1 reading comprehension was the best predictor of L2 reading comprehension, and L1 word decoding was the best predictor of L2 word decoding. The findings suggested that CLT from L1 to L2 for a variety of reading competencies continued for several years after students learned to read and write in L1. It can be said that, differing from some other researchers (e.g., Atwill et al., 2007; Proctor et al., 2006), the results of the study by Sparks et al. (2008) demonstrated the most comprehensive and sustained evidence of CLT in reading.

**Conclusion.** The construct of linguistic interdependence by Cummins (1979) serves as a foundation for the concept of CLT. This concept underpins the relevance of L1 reading instruction for English reading development. There is an extensive body of research on CLT. The extent of CLT was examined with respect to various reading skills: phonemic awareness (Atwill et al., 2007), decoding (Sparks et al., 2003), and reading comprehension (Proctor et al., 2006). The extent of CLT varied from no evidence of transfer (Mathes et al., 2007) to modest (Manis et al., 2004), to comprehensive and sustained (Sparks et al., 2008). The researchers also noted the variability of CLT for different reading skills. For example, according to Proctor et al. (2006), there is a stronger connection between decoding skills in L1
and L2 than between L1 vocabulary knowledge, acting as a proxy for background knowledge, and L2 reading comprehension. In addition, it was established that the extent of CLT may be dependent on the level of language skills in L1 (Atwill et al., 2007; Durgunoğlu et al., 1993). Therefore, ELLs with a low level of language skills in L1 may not benefit from L1 reading to the same extent as children with a higher level of language skills in L1. The role of CLT is not limited to facilitating the development of L2 reading skills: it can also serve as a predictor of L2 reading achievement (Sparks et al., 2008; Lindsey et al., 2003; Gabriele et al., 2009).

**Implications of Existing Research for the Current Study**

Research (e.g., Bernhardt, 2003, Hammer et al., 2011) points to the fact that when ELLs learn to read in English their learning experience is significantly different from that of English-speaking students who learn to read in their primary language. Therefore, when looking for ways to bridge the gap in English reading achievement between ELLs and native speakers, it is necessary to take a broad look at what instructional approaches can help ELLs to become competent English readers.

Some of the literature (e.g., Lentes, 2004; Mathes et al., 2007) on effective reading instruction for ELLs claimed there is insufficient empirical evidence to generate a clear vision of what should be the nature of this instruction. However, one important issue raised in the discussion about what reading instruction can be regarded as optimal for ELLs is the issue of the language of instruction (e.g., Slavin & Cheung, 2005; Domínguez de Ramírez & Shapiro, 2007). By acknowledging that the choice of the language of instruction (L1 vs. L2) may be relevant to the task of ensuring English reading success in ELLs, one assumes a broader look at the issue of L2 reading of ELLs rather than merely viewing it from a monolingual perspective.

The language of instruction has become a topic of substantial controversy in the
education of ELLs (Branum-Martin et al., 2007). While no definitive consensus has been achieved with respect to any aspect of this issue, there is a marked tendency in research (e.g., Slavin & Cheung, 2005; August & Shanahan, 2006) to acknowledge the fact that L1 reading instruction is generally beneficial for ELLs and that bilingual reading programs are more effective in supporting the development of English language literacy than English immersion programs. In addition, research has demonstrated that the benefits of L1 reading instruction tend to outweigh its probable disadvantages (e.g., Branum-Martin et al., 2010; Flood et al., 1997; Rolstad et al., 2005; Slavin & Cheung, 2005).

The potential added value of L1 reading instruction in terms of later English reading achievement of ELLs is supported by studies on CLT (e.g. Atwill et al., 2007; Durgunoğlu et al., 1993; Sparks et al.; 2008; Proctor et al., 2006). While various studies established differing degrees of CLT with respect to a variety of reading skills, it is prudent to assume that CLT can provide some support for L2 reading development. The evidence of CLT agrees with a many researchers’ view that “learning to read in the primary language is a shortcut to reading in a second language” (Krashen, 2002, p.143).

The claim made by Krashen (2002) is in line with the conclusion of the National Literacy Panel on the Language-Minority Children and Youth that “some amount of teaching of students in their home language was beneficial to English literacy learning (better than English immersion)” (August & Shanahan, 2010, p. 342). However, as suggested by August and Shanahan (2010), there are still not enough studies that have examined how to help ELLs to become proficient English readers.

Several limitations of existing research on L1 reading instruction justify the need for the current study. First, there is a tendency to view English reading of ELLs as equivalent to English
reading of monolingual students and to not take into account the factor of bilingualism (Escamilla, 2009). However, “the mere existence of a first language (regardless of whether it is only oral, or oral and literate) renders the second-language reading process considerably different from the first-language reading process” (Bernhardt, 2003, p. 112). An investigation of the effectiveness of bilingual reading programs should account for the L1 oral skills of participating students as well as their developing literacy in L1.

Second, in 1997, Flood et al. believed that the debate about the merits of bilingual education was over due to the ample evidence supporting the benefits of bilingual education and L1 literacy instruction. However, the debate has continued. Moreover, although federal policy neither endorsed nor opposed bilingual education, in recent decades, both federal and state policy changes have discouraged it (Calderón et al., 2011). Therefore, there is a strong need for more research to examine bilingual programs and collect more recent data on their efficacy, particularly in view of the growing ELL population in the United States.

Third, numerous studies and meta-analyses have been conducted with the intent of comparing English immersion and bilingual reading programs for ELLs. The results of these studies have been mixed. For example, a meta-analysis conducted by Slavin and Cheung (2005) established that bilingual programs were preferable over immersion programs in terms of the English reading development of ELLs. Earlier studies by Rossell and Baker (1996) and Thomas and Collier (1997) demonstrated effective results of English immersion with respect to the English language achievement of ELLs. A later study by Slavin et al. (2011) did not support the premise that TBE ultimately helped ELLs to read better in English, though few other studies besides this one have examined TBE programs. The fact that the current study was intended to examine a TBE program provided another compelling reason for undertaking this research.
In addition, multiple studies (e.g., Atwill et al., 2007; Durgunoğlu, 1993; Proctor et al., 2006) have focused on CLT with respect to various individual reading skills without examining a comprehensive set of skills that allows ELLs to become proficient English readers. The approach taken by these studies sets a limit to a more comprehensive evaluation of CLT from L1 to L2. Therefore, there is still much to be learned about how reading ability in L1 translates into overall reading achievement in L2. Finally, this review uncovered a clear gap in the literature on how the methodology used for placing ELLs in L1 versus English-only reading instruction may influence later L2 reading achievement in elementary ELLs.

Regardless of these limitations, based on the analysis of the reviewed literature the question about the connection between L1 instruction and English reading achievement in Spanish-speaking ELLs suggests the following thesis: although there is a tendency to view L2 reading development from a monolingual perspective and there remains a lack of consensus on effective reading instruction for ELLs in general and with respect to the language of instruction in particular, the evidence in support of CLT from L1 to L2 suggests that L1 reading instruction in bilingual programs can have a beneficial effect on the English reading development of ELLs.

**Chapter Summary**

The thesis of the study can be summarized using a declaration by Krashen (2002): “In early stages, reading in the first language is of great help in promoting second-language literacy” (p. 148). This thesis includes a number of key ideas.

First, when seeking effective instructional approaches for ensuring English reading achievement in ELLs it is important to pay close attention to the language of instruction (Nakamoto et al., 2012). The topic of the language of instruction has been controversial and
highly debated in the education field in the U.S. Regardless of the ongoing debate, research suggests that the use of students’ L1 for instructional purposes supports the learning processes in ELLs (e.g., Rolstad et al., 2005; Slavin & Cheung, 2005). However, a positive effect of L1 reading instruction can be influenced by the level of language proficiency in L1 (e.g., Atwill et al., 2007; Durgunoğlu et al., 1993).

Second, although there is a need for more quality research that examines the achievement outcomes of bilingual and immersion programs, meta-analyses of existing studies supported the premise that there are more positive effects on English literacy in bilingual reading programs than in English immersion programs (Rolstad et al., 2005; Slavin & Cheung, 2005).

Third, the existence of the positive effect L1 reading instruction can have on L2 reading is supported by extensive research on the CLT of various reading skills from L1 to L2 (e.g., Atwill et al., 2007; Gabriele et al., 2009). This evidence buttresses the idea that “first-language reading skills are related to second-language reading skills” (August et al., 2006, p. 362) and that there is a connection between L1 literacy and L2 literacy with respect to ELLs (Escamilla, 2009).

The foregoing points have clear and direct implications for research and practice. In terms of research, the thesis emphasizes the relevance of the issue of L1 reading instruction to L2 reading proficiency. Therefore, this research may encourage more empirical research with respect to best practices for the reading instruction of ELLs.

In practical terms, the thesis statement provides guidance in addressing a critical issue of the gap in English reading proficiency between ELLs and monolingual students. In the current climate of growing school accountability, standardized measures are becoming increasingly relevant. However, the literature review clearly indicated the paucity of research on the effect of L1 reading instruction on the overall English reading achievement of ELLs as demonstrated on
standardized measures of reading. There is also a lack of research on how different selection criteria for placing ELLs either in L1 or English-only reading instruction upon entering kindergarten may influence their later English reading achievement. At the same time, many schools in the United States have been unable to meet their accountability requirements due to the underperformance of ELLs in reading (Slavin & Cheung, 2005). More importantly, ELLs need adequate English reading skills to meet increasing academic requirements. Ultimately, American society cannot provide equal opportunity for all its members if schools are not successful in assisting a growing population of ELLs to achieve the level of academic success associated with more opportunities in education and in life in general (Slavin & Cheung, 2005).
CHAPTER 3: METHODS AND PROCEDURES

Methodology

Fraenkel, Wallen and Hyun (2011) stated that quantitative research studies are aimed at exploring relationships between variables. The present research study examined the effects of a change in the screening procedure used for placing ELLs either in L1 or English-only reading instruction on their English reading achievement as measured by standardized reading test scores. The study sought to determine whether there is a significant relationship between a change in the screening procedure for ELLs and their later English reading achievement. In addition to establishing a relationship between variables, quantitative research seeks to determine the cause of this relationship (Fraenkel et al., 2011), that is, to establish a cause and effect, also known as causality. Muijs (2011) argued that quantitative methods are best for establishing causality. This study investigated causality between a change in the screening procedure used for placing ELLs either in L1 reading instruction or English-only reading instruction and later English reading achievement of these ELLs. Shadish et al. (2002) emphasized that while the nature of the causal knowledge is localized to a specific study researchers aspire for more generalized causal goals, that is, to establish dependable causal relationships that can be used by practitioners, policy-makers, and other researchers. This statement by Shadish et al. (2002) is supported by Muijs (2011) who argued that quantitative research studies commonly seek to generalize from a specific sample to the population and that various statistical methods employed in quantitative research allow for such generalizations. This research study also made an effort to generalize findings from a specific sample to the population. Accordingly, a quantitative methodology was an appropriate choice of methodology for the study.
Fraenkel et al. (2011) pointed out that “quantitative research has established a widely agreed-on formulation of steps that guide researchers in their work” (p. 10). These steps include selecting an appropriate research design, sampling method, and instrumentation as well as defining procedures and a data analysis plan. The above steps and the rationale for choosing a specific research design and research techniques will be discussed in this chapter.

**Research Design**

The study used a quasi-experimental research design. A quasi-experimental design is defined by Creswell (2012) as a between-group experimental design. According to Creswell (2012), the between-group design in which the researcher compares two or more groups is widely used in education research. True experiments are commonly not feasible in an education setting where parents may be opposed to their children’s participation in research and where most educators do not have the sufficient degree of control required for conducting a true experiment (Hoy, 2010). Additionally, quasi-experiments are appropriate for research situations where it is not possible to create artificial groups and, for that reason, conduct a true experiment (Creswell, 2012). Therefore, in a natural school setting, quasi-experiments are intended “to approximate as closely as possible the advantages of true experimental designs” (Muijs, 2011, p. 23). Moreover, the fact that the goal of this study was to examine the effect of an intervention (change in screening procedure used to place ELLs either in L1 or English-only reading instruction) that was implemented before this research study was conceived provided sufficient rationale for the choice of a quasi-experiment as the research design for this investigation.

When random assignment is not possible, a quasi-experiment where participants are assigned to groups non-randomly is a second-best alternative for making causal inferences (Vogt, 2007). With respect to this study, the primary advantage of the quasi-experiment is that it
allowed for the comparison of student performance before and after the change in screening procedure first with respect to all ELLs and then with respect to the two following groups of students: (1) ELLs who received L1 reading instruction and (2) ELLs who received English-only reading instruction. The research questions formulated for the study called for establishing a causal effect of the independent variable (type of screening procedure) on the dependent variable (English reading achievement test scores).

At the same time, a quasi-experimental design has certain disadvantages. In a real school environment, such as the research site for this study, while trying to make comparison and experimental groups as similar as possible, it is not feasible to totally equalize factors such as family socioeconomic status, school preparedness upon entry in kindergarten, family ability and willingness to provide academic support, gender, and student ability (Muijs, 2011). Teacher quality, school culture, and peer group effects constitute another set of factors that may have impacted the outcomes of this quasi-experimental study (Muijs, 2011). Any and all of the above factors may have influenced the results of this research and, therefore, needed to be taken into account when making conclusions about the research findings. In addition, as this study was conducted at an elementary school in the Pacific Northwest, the ability to generalize the findings to the entire population of Spanish-speaking ELLs in other schools across the country may be limited (Vogt, 2007). However, whenever a true experiment is not possible and/or not realistic “most researchers would agree that quasi-experimental evidence is better than no evidence at all” (Vogt, 2007, p. 108).

Spanish-speaking ELLs were the research participants in this study. As shown in Table 2, the participants belong to six cohorts of students. Each cohort consists of two groups: (1) ELLs who received initial L1 reading instruction and (2) ELLs who received English-only (L2) reading instruction.
Table 2

*Formation of Pre-And Post-Intervention Student Cohorts*

<table>
<thead>
<tr>
<th>Pre-intervention</th>
<th>Post-intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort 1</td>
<td></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td></td>
</tr>
<tr>
<td>Cohort 2</td>
<td></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td></td>
</tr>
<tr>
<td>Cohort 3</td>
<td></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td></td>
</tr>
</tbody>
</table>

All cohorts were formed upon students’ enrollment in kindergarten followed by the identification of students as ELLs. In each cohort, ELLs were placed in L1 or English-only reading instruction on the basis of language proficiency screening. The Woodcock-Muñoz Language Survey (Woodcock & Muñoz-Sandoval, 1993b, 1993c) was administered in the English (LS-E) prior to the intervention and both in the English (LS-E) and the Spanish (LS-S) versions after the intervention. The survey is primarily designed to measure academic language proficiency (CALP) in both languages as well as an overall measure of language competence. The survey offers a procedure to establish cut points for five levels of CALP in both languages: (1) negligible, (2) very limited, (3) limited, (4) fluent, and (5) advanced. The Woodcock-Muñoz Language Survey can be used for a wide range of purposes which include classifying a subject’s language proficiency, determining eligibility for bilingual services, and describing the language characteristics of subjects in research studies. The survey includes four tests each of which
measures a different aspect of language proficiency. For interpretive purposes, four tests are also combined in clusters or broad-based measures: broad language ability cluster, oral language cluster, and reading-writing cluster. The Woodcock-Muñoz Language Survey is a norm-referenced standardized measure, and the information supporting its validity and reliability will be provided later in the chapter.

During the three pre-intervention school years (2004/5, 2005/6, and 2006/7) ELLs were placed in the groups receiving L1 reading instruction and English-only reading instruction on the basis of their English language proficiency upon entry to kindergarten. Accordingly, students who scored 1 and 2 on the Woodcock-Muñoz Language Survey-English (LS-E) were assigned to the group receiving L1 reading instruction. Students who scored 3 and 4 on the Woodcock-Muñoz Language Survey-English (LS-E) were assigned to the group receiving English-only reading instruction. The level of Spanish language proficiency was not established or taken into account for placement purposes.

Following an intervention at the beginning of the 2007/8 school year, the criteria for student placement was changed. The rationale for the change originated from the Cummins (1979) threshold hypothesis which considers the level of L1 proficiency as a threshold required to activate the principle of linguistic interdependence and to allow for consequent cross-linguistic transfer of skills from L1 to L2. In line with the new placement criteria, ELLs were screened both for English (L2) and Spanish (L1) proficiency with the use of the Woodcock-Muñoz Language Survey-English (LS-E) and the Woodcock-Muñoz Language Survey-Spanish (LS-S) and the level of L1 proficiency was regarded as the main selection criterion for placement. Accordingly, ELLs scoring 3 and 4 in Spanish were placed in the group receiving L1 reading instruction, and ELLs scoring 1 and 2 in Spanish were placed in the group receiving English-only reading instruction.
As a result, ELLs with higher L1 proficiency received L1 reading instruction rather than English-only reading instruction. This new approach engenders a greater focus on students’ stronger language. For ELLs whose L1 is stronger, this approach may promote later CLT from L1 to L2. Ultimately, by capitalizing on students’ strong L1 skills the new selection criteria were aimed at helping ELLs “to avoid the failure experience that is likely if they are asked to learn both oral English and English reading at the same time” (Slavin et al., 2011, p. 48).

In third grade, all ELLs in the six cohorts took the OAKS standardized English reading achievement test mandated by ODE. By that time, all ELLs received four years of reading instruction in a respective group: L1 reading instruction or English-only reading instruction. The annual administration of the OAKS test during six consecutive years created six time points when the data used for the purpose of this research study were generated. Table 3 provides a graphic representation of the data generation time points for each respective student cohort.

Table 3

**Data Generation Time Points**

<table>
<thead>
<tr>
<th>Pre-intervention</th>
<th>Post-intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cohort 1</strong></td>
<td><strong>Cohort 1</strong></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td>(L1 and L2 groups)</td>
</tr>
<tr>
<td><strong>Cohort 2</strong></td>
<td><strong>Cohort 2</strong></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td>(L1 and L2 groups)</td>
</tr>
<tr>
<td><strong>Cohort 3</strong></td>
<td><strong>Cohort 3</strong></td>
</tr>
<tr>
<td>(L1 and L2 groups)</td>
<td>(L1 and L2 groups)</td>
</tr>
</tbody>
</table>
Sampling Methods

Research site. An elementary school in the Northwest was the research site for the study. The current enrollment at the school is about 650 students in grades K through 5. The majority of the students come from families with low socioeconomic status. Accordingly, about 80% of the students qualify for free and reduced cost lunch. In addition, approximately 51% of the student population qualify for ELL status and receive English-as-a-second-language instruction on a pull-out basis. The population of ELLs is highly homogenous: all students are native Spanish speakers. The school has been operating the Native Language Literacy (NLL) program in grades K through 3 to support the English reading development of ELLs since 2004.

The school was identified as the research site for the study because of the availability of “well-defined, intact groups that are easily studied” (Creswell, 2012, p. 323) and consisted of potential participants, that is, ELLs who either received or did not receive initial L1 reading instruction. The school has been using L1 reading instruction to support the English reading achievement of ELLs since 2004. As a result, there was a significant amount of data (standardized English reading achievement test scores for both categories of ELLs) available for the purpose of this research. The combination of these factors made the school a suitable research site for studying the effect of a change in the screening procedure used to place ELLs either in initial L1 or English-only reading instruction on their later English reading achievement.

Participants. The research participants in the study were Spanish-speaking ELLs with varying degrees of English language proficiency and currently in grades 5 through 10. According to the requirements of ODE, students were identified as ELLs on the basis of parental responses to the Home Language Survey at the time of enrollment. All students were identified by their parents as Hispanic. The ratio of males/females was respectively 49% and 51.
students come predominantly from families with low socioeconomic status (about 80% of the entire student population are eligible for free or reduced lunch) and low levels of parental education attainment. Access to the participants was gained through the researcher’s employment at the school in a full-time teaching position.

The study used convenience sampling strategy. The total projected available sample was 307 Spanish-speaking ELLs who are currently in grades 5 through 10. These students were identified as ELLs at the time they entered kindergarten. Three pre-intervention cohorts consisted of 161 students. Three post-intervention cohorts consisted of 146 students. The sample was disaggregated according to the language of reading instruction: the pre-intervention group that received L1 reading instruction consisted of 58 students, and the pre-intervention group that received L2 reading instruction included 103 students. Accordingly, the post-intervention group of students who received L1 reading instruction consisted of 56 students, and the post-intervention group who received L2 reading instruction included 90 students.

However, to ensure the validity of the study some students had to be excluded from the total projected sample. The rationale and the criteria used for the exclusion are discussed in detail in the Sample Characteristics section of Chapter 4. Accordingly, the adjusted sample for the study consisted of 245 ELLs with 128 ELLs in the pre-intervention group and 117 ELLs in the post-intervention group. The overall L1 instruction group consisted of 98 students, and the overall English-only (L2) instruction group consisted of 147 students.

The convenience sampling method was used due to the fact that the researcher’s employment provided easy access to both the site and participant data that were studied to address the research questions and test the research hypotheses. According to Muijs (2011), convenience sampling is “the most common sampling method in education studies at present” (p. 36).
Convenience sampling can provide useful information for answering research questions and testing hypotheses (Creswell, 2012). While the research results can never be expected to be generalizable from a specific sample to the general population with 100% certainty (Muijs, 2011), it is important to take steps to address the threats to generalizability resulting from the weaknesses in sampling. First, researchers should exercise extreme caution when generalizing results from convenience samples (Vogt, 2007). Second, it is important to bear in mind the level of generalization (Vogt, 2007). While cautious generalization of the study findings to the school level is appropriate, broader generalizations to other populations of ELLs at other sites should not be regarded as feasible and would require a replication of the study. Third, due to the fact that “the best the researchers can do is to assess generalizations in terms of the weight of the evidence at a given point in time” (Hoy, 2010, p. 19), it is important to provide sufficiently detailed information that supports the dependability of any generalizations made in the study. Detailed information is also needed to allow for study replication with respect to different populations and/or settings. Every effort has been made to provide such information within the framework of this study.

In the course of the study, no information was obtained directly from the participants, and no action was required on the part of the participants. Therefore, no steps had to be taken either to motivate or to induce the participants to become involved in the study. Although as an ESL and reading teacher, the researcher previously came in contact with many ELLs at the school, no research relationship was established between the participants and the researcher within the framework of the study.

**Procedures**

The first step in the data collection process for this study was to seek approval from the district to access and use the data for the purpose of this research. The next step was to obtain a
complete list of the target population (ELLs) from the student information system at the research site (elementary school). Next, it was necessary to obtain third grade reading test scores of ELLs for the relevant school years from the existing data sets of the OAKS test scores at the ODE website. Then, the target population was divided into two groups: pre-intervention and post-intervention. Finally, each group was divided into two subsets according to the independent variable (language of reading instruction): L1 and L2 reading instruction.

Prior to entering and analyzing the data the researcher obtained an approval from the Northeastern University Institutional Review Board by filing an application for Approval for Use of Human Participants in Research. This application contained information about the goals and purpose of the study, described recruitment and study procedures, addressed potential risk and benefits for participants and outlined procedures to be used in order to ensure anonymity of data during collection and entry of data. The application was supported with a permission letter from the school district, an Assurance of Principal Investigator form, and a Certificate of Completion of Protecting Human Research Participants training course.

**Instrumentation**

The data collection instrument employed in the study was the third grade standardized English reading achievement OAKS test. The OAKS is an element of the Oregon Statewide Assessment System (OSAS) developed by the ODE. The OAKS measures student performance in mathematics, reading, science, and social studies via multiple-choice tests aligned to grade level content standards. The tests are administered in grades 3 to 8 and grade 10. The OAKS is a summative assessment. Summative assessments are typically used to assess student achievement levels.

The OAKS is administered in an online format (OAKS Online) and is an adaptive
assessment. The computer program selects items varying in level of difficulty depending upon the student’s performance on the previous item. This means there is a grade-level pool of items rather than a single set of items for each grade. Thus, the test encourages students to perform at their highest ability. Each test item measures a single score reporting category (SRC) or strand. The test pool contains items of varying levels of difficulty for each SRC. Typical items have four response options. The entire test generally consists of 40 to 50 items. The third grade reading test covers the following strands: vocabulary, read to perform a task, demonstrate general understanding, and develop interpretation. Sample items from third grade reading achievement test for each SRC are presented below:

Vocabulary:

From the way it is talked about in the story you can tell that a brook is most like

1 a tree.
2 a stream.
3 a thorn.
4 a tongue.

Develop Interpretation:

Why did Jessie whisper when she said, “Keep still!” in this story?

1 She didn’t want to wake up Violet.
2 She didn’t want them to scare the dog.
3 She didn’t want to bother the neighbors.
4 She didn’t want what was in the bushes to hear them.

Demonstrate General Understanding:

Who is telling this story?
Reading to Perform a Task:

If you traveled to Thailand in October, which festival might you see?

1. Flower Festival
2. Candle Festival
3. Thai New Year
4. Rocket Festival

The dependent variable stipulated in the research questions is English language reading achievement of ELLs. One of the primary purposes of the OAKS is to ascertain the achievement level of individual students. Therefore, the standardized reading test used within the framework of OAKS is a logical and appropriate instrument to measure the reading achievement of ELLs (those who received L1 reading instruction in grades K through 3 and those who received English-only reading instruction in the same grades) before and after the intervention. By comparing the measures of the English reading achievement of ELLs before and after the intervention this study strove to determine the effect of the change in the screening procedure for placing ELLs either in L1 reading instruction or English-only reading instruction on their English reading achievement.

The OAKS is mandatory for all public schools in the state. The research site for the study is located in Oregon, and, therefore, all the participants in the study took third grade achievement tests as required by ODE. The test scores are accessible to the teachers and school administrators at the ODE website. However, the researcher had to obtain the permission of the
district superintendent to use third grade reading test scores of ELLs who took the test during the school years relevant to the research study.

According to Fraenkel et al. (2011), validity and reliability are critically important factors that should be carefully considered when selecting a research instrument. Fraenkel et al. (2011) attributed the importance of instrument validity to the fact that researchers need to be confident that the data collected with the use of the selected instrument can serve the purpose of the research study. The technical adequacy of an assessment is supported through the reliability, content validity, criterion validity, and construct validity of its tests scores. The information on the components of technical quality of the OAKS reading test is provided by ODE (2007) in the Oregon Statewide Assessment System Annual Technical Report.

Muijs (2011) described instrument validity as a multidimensional construct and pointed out that a valid instrument should sufficiently demonstrate all three types of validity: content-related, criterion-related, and construct-related. Fraenkel et al. (2011) defined content validity as the extent to which the content and the format of the instrument are consistent with the variable. This type validity is normally determined through expert judgment. Content-related evidence of validity of the OAKS reading test includes the following: (1) the test is based on the rigorous content standards of student knowledge; these standards were developed and reviewed by state educators; (2) test specifications establish a clear and direct link between test content, content standards, and performance levels; (3) test items were developed by panels of educators through a consensus-driven development process; and, (4) ongoing studies were conducted to ensure and increase alignment among content standards, instruction, and assessment.

Fraenkel et al. (2011) stated that criterion validity is the degree to which performance on an instrument is related to performance on other instruments meant to measure the same variable
or other variables that are related to the measured variable. There are two types of criterion validity: predictive and concurrent. Concurrent validity is usually established by comparing the performance on the instrument that is being validated with the performance on other independent criteria, such as other instruments used to measure the same independent variable (Fraenkel et al., 2011). The key index of criterion-related evidence is the correlation coefficient. The correlation coefficients of the OAKS reading test demonstrate that the scores are highly correlated with scores on the California Achievement Tests (.75), the Iowa Test of Basic Skills (ITBS) (.78), the Northwest Evaluation Association (NWEA) subject tests (.73), and the Lexile Scale (.77). Predictive validity is related to the outcomes that the instrument can theoretically predict (Muijs, 2011). With respect to the OAKS, there is a clear association between the scores of students who achieved proficiency levels (performance standards in content areas) on the OAKS test and their later performance during the first year of college. These students demonstrated higher GPA, higher SAT scores, and higher retention.

Construct validity is the degree to which the instrument measures a hypothetical intended construct (Fraenkel et al., 2011), such as student reading achievement in the case of this study. Fraenkel et al. (2011) posited that construct validity is the most complex of the three types of validity and that there is no single type of evidence that can support it. Construct validity is related to the significance, meaning, and use of scores from an instrument (Creswell, 2012). Therefore, construct validation can include a variety of procedures and be supported by criterion-related and content-related types of evidence. The construct validity of the OAKS test is based on the evidence from tests specifications and test development procedures. Test specifications demonstrate a clear link among the test content, the content standards, and corresponding performance levels. “Test specifications place
boundaries around the domains for which score inferences are desired” (ODE, 2007, p.13). In addition, test specifications include appropriate types of items related to academic content and are linked to Performance Level Descriptors (PLD) designed to help educators understand how the academic standards are related to student performance at different achievement levels. “Test item faithfulness to the Academic Content Standards is strengthened by using subject matter experts to design the test frameworks” (ODE, 2007, p.14). Expert judgment is also a source of evidence supporting the assumption that test items adequately reflect the desired content domains. In addition, expert judgment provided information about potential threats to validity, such as construct-irrelevant elements in the test items. The abovementioned content-related and criterion-related evidence and independent expert judgment support the construct validity of the OAKS test with respect to “the internal structure of an instrument and the concept it is measuring “(Muijs, 2011).

According to Fraenkel et al. (2011), reliability is determined by the degree to which test scores consistently measure the variable the instrument is supposed to measure. The evidence of the instrument reliability is derived from the analysis of the standard error of measurement (SEMeas). Fraenkel et al. (2011) described SEMeas as an index that estimates the degree to which a score would change under different circumstances. The average SEMeas for third grade reading tests was calculated at 3.2 Rasch Unit (RIT) points. The RIT scale is a measurement scale used to measure student achievement and growth. This measurement scale was developed to simplify the interpretation of test scores. The SEMeas analysis suggests that the OAKS provides similarly reliable tests scores across the ability range with the exception of the extreme ends of the distribution, that is, very high and very low scores. However, the existing data demonstrated that most ELLs did not score at the extreme ends of the distribution.
In The Oregon Statewide Assessment System Annual Technical Report (ODE, 2007) providing information on validity and reliability, the results of the SEMeas analysis were described as the measure of reliability of the OAKS third grade reading test. However, it should be noted that there are two other methods that are commonly used to answer the question about test reliability also referred to as internal consistency: split-half reliability and Cronbach’s Alpha. However, according to the Northwest Evaluation Association (NWEA) (2004), these two methods have two limitations with respect to the estimates of internal consistency: (1) their results are sample dependent, that is, these measures are based on student responses to the test; therefore, different samples may result in different estimates of internal consistency; (2) these estimates are based on the assumption that all students take a test consisting of the same items.

In view of the foregoing, NWEA (2004) suggested that another method to estimate internal consistency should be used. This method is known as the marginal reliability coefficient and is based on the underlying scale (the RIT scale in the case of the OAKS test) and “test information” as two test characteristics. This method uses measurement error estimated at different points on the achievement scale combined into a single index. The measurement error will always be minimal for the section of the underlying scale in the middle of the test and will increase toward the ends of the scale. In addition, this method does not require test administration. Finally, NWEA (2004) stipulated that this method provides estimates that are nearly identical to Cronbach’s alpha.

The OAKS reading test was selected as the instrument for the study for a number of reasons. First, the test has proven validity and reliability supported by research studies (ODE, 2007). Second, the test provides a comprehensive snapshot of student English reading achievement (as defined by grade level content standards) at a critical milestone in the literacy
development: by the end of third grade, students are expected to complete their learning to read and begin applying their reading skills to learn content. Fourth, the administration of the test approximately coincides with the completion of L1 reading instruction. Next, all students take the test, and the historical data are available. Finally, as a subject instrument (Fraenkel et al., 2011), the test is directly related to the dependent variable (English reading achievement).

Data Analysis Plan

Preparation of the data. The first step in analyzing quantitative data is preparing data for analysis. This process involves several steps: scoring the data, determining the type of scores to analyze, selecting a statistical program and inputting the data into the program, and cleaning the data for analysis (Creswell, 2012). The data in the study are represented by student scores on a standardized reading achievement test. Therefore, there was no need to score the data.

The next step was to input the data. This process involved importing the scores from the OAKS database to a spread sheet. After the data were input, it was necessary to code them before the data were handed over to a statistical consultant to conduct statistical analyses. Data coding served the purpose of de-identification of the data and thereby ensured data anonymity in order to maintain confidentiality. Confidentiality was necessary to protect human subjects participating in the proposed study. Participants in the study were de-identified by removing their names and Secure Student IDs (SSID) and assigning each participant a code. Each code reflects the following parameters: (1) pre- or post-intervention group, (2) language of reading instruction, (2) observation year, and (4) number on the subgroup list.

Although the student population at the school is fairly stable, it was important to take into account the fact that there may have been some ELLs who took third grade English reading achievement test but did not receive reading (either L1 or L2) instruction at the school in grades
K through 3. This situation appears to be the opposite of attrition defined by Shadish et al. (2002) as a loss of response from participants after they were assigned to conditions. However, similar to attrition, the inclusion of the test scores of such “additional” students in the data for analysis could compromise internal validity of the study because these ELLs (1) may have received different types of reading instruction; (2) were instructed by different teachers; and, (3) most importantly, were not screened for language proficiency upon entry to kindergarten to be placed either in L1 or English-only reading instruction. To control for these ELLs, it was necessary to use historical enrollment data for the school years from 2004 to 2013 and verify the data by excluding such ELLs from the pool of participants. As the data in this study are represented by the scores on the standardized reading achievement test there was no need to transform, recode or collapse the data to conduct data analyses and test the hypotheses.

**Choice of statistical technique.** Once the data were entered into a statistical analysis program, it was necessary to select statistical methods appropriate for the research design, the type of variables, the research questions, and the hypotheses for this research study. As the dependent variable scores came from six observations, the analyses to be performed for this research study fall under the category of short or abbreviated interrupted time series (AITS) analysis (Wong, Cook, & Steiner, 2009).

Shadish et al. (2002) argued that the interrupted time series can be regarded as one of the most effective and powerful quasi-experimental analyses. The interrupted time series is a specific type of time series used to assess treatment impact (Shadish et al, 2002). According to Shadish et al. (2002), the interrupted time series should be applied only to situations where the exact point when a treatment occurred can be clearly identified. The fact that the point at which the treatment (the change in the screening procedure) occurred was clearly identifiable justified
the choice of the interrupted time series for this research. The choice of the interrupted time series was also supported by the fact that Shadish et al. (2002) viewed the interrupted time series as “a particularly strong quasi-experimental alternative to randomized designs when the latter are not feasible and when a time series can be found” (p. 172). The annual English reading achievement test scores can be regarded as observations made repeatedly over six years. According to Tabachnick and Fidell (2007), the availability of data from observations repeated over a period of time allows for the use of the time series analysis.

Box, Jenkins, and Reinsel (1994) maintained that it was necessary to have a large (up to 100) number of observations to identify a stable model in an interrupted time series. Contrary to Box et al. (1994), Shadish et al. (2002) argued that an abbreviated time series can be sufficient for identifying the model. This position was supported by Wong, Cook, and Steiner (2009) who pointed out that the interrupted time series design has had little use in education research due to the fact that collecting the time series data with a large number of observations in an education setting can be highly problematic. Wong et al. (2009) argued that the requirement for a large number of observations, as stipulated by Box and Jenkins (1970), is inapplicable to most research studies in education and suggested that the abbreviated interrupted time series (AITS) should be used instead. According to Wong et al. (2009), the rationale for the use of AITS can be justified by the fact that longitudinal research studies in education typically collect data only at three or four points before an intervention and at even fewer points after an intervention. For the purposes of this study, the data were available at three time points before the intervention and at three time points after the intervention. This renders the short time series analysis a suitable approach for estimating the causal effects of the change in the screening procedure.
**Data analysis.** Campbell and Stanley (1966) who originally proposed the use of time series diagrammed the interrupted time series as follows:

```
......O O O O O O X O O O O O O O......
```

In this diagram, O represents an observation, and X represents an intervention. An intervention “interrupts” a time series by dividing it into two discrete segments: pre-intervention and post-intervention. As a result, the analysis of the interrupted time series data is usually focused on the magnitude and the form of the impact of an intervention on the time series (McDowal, McCleary, Meidinger, & Hay, 1980). Therefore, according to McDowal et al. (1980), the interrupted time series data can be analyzed by means of the statistical comparison of the data in pre-intervention and post-intervention segments.

Shadish et al. (2002) noted that ordinary statistics should not be used to analyze the data in the interrupted time series. For example, according to Shadish et al. (2002), a t-test cannot be used to compare pre-intervention observations with post-intervention observations. This position is supported by Box and Tiao (1975) who maintained that an evaluation of t for comparing means is irrelevant with respect to analyzing interrupted time series. More recently, Lagarde (2011) also cautioned against a simplistic approach to the analysis of longitudinal data in interrupted time series. According to Lagarde (2011), although there may not be an opportunity to establish a baseline before an intervention, routine data collected during a pre-intervention period can still be available. Lagarde (2011) argued that the use of these data allows researchers to evaluate the impact of interventions implemented at a precise time point. However, Lagarde (2011) stipulated that analyzing longitudinal data through a simple comparison of mean outcomes before and after an intervention is not only limiting but, in addition, may be misleading due to producing biased results and overestimating the impact of an intervention. Lagarde (2011) listed non-stationarity,
auto-correlation, and seasonality as other potential sources of bias that may result from simple comparison of means when analyzing longitudinal data in time series. Non-stationarity implies that the time series data normally has a natural trend, e.g., an upward slope. Auto-correlation is related to the fact that ordinary statistical methods are based on the assumption that errors associated with each observation are uncorrelated. Failure to account for this when analyzing time series data may lead to incorrect results with respect to statistical significance of coefficient. Finally, time series data may experience “seasonality” or periodical fluctuations independent of other factors. Failure to control for seasonality may distort the real effect of an intervention.

According to Lagarde (2011), Auto Regressive Integrated Moving Average (ARIMA) modelling developed by Box and Jenkins (1970) has been a primary method of analyzing data in interrupted time series. Huitema, McKean, and Laraway (2007) pointed out the ARIMA modelling has been endorsed by many leading methodologists (e.g., Glass, Wilson, & Gottman, 1975; McCleary & Hay, 1980). However, other authors (e.g., Gorman & Allison, 1997) argued that ARIMA models have certain features that may make their use problematic. First, ARIMA models require a large number of observations (at least 100 points) (Lagarde, 2011). This fact complicates the use of ARIMA models in healthcare (Lagarde, 2011) and behavioral research (Huitema et al, 2007). The situation may be equally applicable to education research where it is often not possible to collect data at so many points. Next, Lagarde (2011) pointed out that, due to the fact that ARIMA modelling originated in econometrics, it has been primarily used to forecast rather than to explain outcomes of interest. Finally, due to its analytical complexity, ARIMA modelling requires highly sophisticated statistical skills on the part of the researcher (Lagarde, 2011; Huitema et al., 2007).

The abovementioned concerns with respect to ARIMA modelling prompted the
development of alternative methods for analyzing interrupted time series data (Huijema et al., 2007). Lagarde (2011) posited that another approach that can reduce difficulties in analyzing time series data is linear regression. Other authors also referred to this method as segmented regression (e.g., Gillings, Makuc, & Siegel, 1981; Wagner, Soumerai, Zhang, & Ross-Degnan, 2002). According to Wagner et al. (2002), segmented regression analysis is “a powerful statistical method for estimating interventions effects in interrupted time series studies” (p. 299) because it allows for statistical evaluation of how an intervention changed an outcome of interest. Wagner et al. (2002) regarded segmented regression analysis as an appropriate method to evaluate the effects of various types of interventions in interrupted time series with the data on continuous or counted outcome measures collected over time at evenly spaced intervals. When interrupted time series are analyzed using segmented regression the estimation of the intervention effect is made by comparing the trend in the outcome after an intervention to the trend that existed in the pre-intervention period (Taljaard, McKenzie, Ramsay, & Grimshaw, 2014). In addition, the use of segmented regression ensures control for baseline (Zhen-qiang, Kuller, Fisher, & Ostroff, 2013).

The data in this study represented by student scores on a standardized reading achievement test can be classified as longitudinal data routinely collected at evenly spaced intervals (annually during six consecutive years). The intervention (change in the screening procedure for placing kindergarten ELLs either in L1 or English-only reading instruction) took place after the third data collection point. Therefore, it is appropriate to treat the data as an interrupted time series. The research questions posed by this study called for examining an effect the intervention may have had on the outcome of interest, that is, English reading achievement of ELLs as demonstrated by their scores on a third grade standardized reading
achievement test. Therefore, the study was not aimed at predicting the outcome of interest but was rather focused on examining the outcome that had already taken place. Finally, the data in the interrupted time series available for the purpose of the study had a limited number of data collection points or observations (three pre-intervention points and three post-intervention points). All of the above aspects of the current research warranted the use of segmented regression, as opposed to ARIMA, as a statistical technique most appropriate for the data analysis in this study.

Validity, Reliability, and Generalizability

The validity and reliability of this study needed to be examined from two perspectives: validity and reliability of the research design and validity and reliability of the instruments. The validity and reliability of the research design are reviewed as applicable to the quasi-experimental research design and the interrupted time series analysis. The validity and the reliability of the data collection instrument (OAKS English reading achievement test) were addressed in the Instrumentation section of Chapter 3. In addition to the data collection instrument, there is another instrument germane to the study: the Woodcock-Muñoz Language Survey used for screening English and Spanish-language proficiency of ELLs. The validity and reliability of this language proficiency measure is pertinent to the research study due to the fact that kindergarten ELLs were placed either in L1 or English-only reading instruction based on the levels of their language proficiency as determined through the administration of this measure. Accordingly, the validity and reliability of the language survey are addressed later in this chapter after the review of the validity and reliability of the study design. The chapter is concluded with the discussion of generalizability as applicable to this study.
**Validity of the research design.** Shadish et al. (2002) observed that in the case of the interrupted time series, history is the primary threat to internal validity. According to Shadish et al. (2002), this threat manifests itself in the possibility of another influence on the dependent variable at the same time of the intervention. Shadish et al. (2002) stated that several controls for the history threat are possible: (1) a no-treatment time series, (2) monthly rather than yearly intervals, and (3) the list of effect-causing events that could have influenced the outcomes. With respect to this study, the history threat should have been controlled as there have been no other changes in the reading instruction of ELLs throughout the years for which data have been collected. The same reading program was used for both English and Spanish reading instruction throughout the nine years included in the study. While there was some limited teaching staff turnover with respect to teachers delivering reading instruction in grades K through 3, the effort of the school administration to maintain the consistency and continuity of the schoolwide approach to reading instruction through the use of the same reading program and the same instructional strategies across all grades should have helped to minimize the teacher effects. Within the framework of this study, it was not possible to have a no-treatment time series or change the intervals to monthly rather than yearly. The list of possible effect-causing events can include changes in student demographics and changes in reading instruction. The fact that no such changes occurred helped to reduce the history threat.

According to Shadish et al. (2002), another threat to validity is instrumentation. The research data (student test scores on the standardized English reading achievement test) used in the study were collected during annual observations. During this period of time, there were no changes to the data collection instrument (OAKS third grade standardized English reading achievement test). Therefore, this consistency of data collection and data storage should have
minimized the instrumentation threat to validity.

Shadish et al. (2002) pointed out that selection can be yet another threat to validity. This threat is also referred to as the subject characteristics threat (Fraenkel et al., 2011). The subject characteristics threat means that the groups may not be equivalent on some important variables other than the identified group membership variable (Fraenkel et al., 2011). Bloom (2003) argued that in the interrupted time series analysis it is necessary “to control explicitly for systematic differences over time in the background characteristics of student cohorts” (p. 9) using school records on student characteristics. During the time frame of the study, the ELL population at the school remained sufficiently homogenous and stable on a number of essential characteristics. The first characteristic is family socioeconomic status. The demographic data on the ELL population at the school show that ELLs predominantly come from families with low socioeconomic status. The ethnicity of the participants is 100% homogeneous: all ELLs were identified as Hispanic and native Spanish speakers. The demographic data demonstrated the male/female ratio close to 50/50. While historically there has always been a moderate level of attrition in the student population, the school did not experience any significant changes in student characteristics during the time frame of the study. Therefore, it can be concluded that selection should not have presented a significant threat to internal validity.

Shadish et al. (2002) pointed out that “the typical statistical conclusion validity threats apply as much to time series as to any other design” (p. 180). These threats include low power, violated test assumptions, and unreliability of measurement. Reliability of measurement means that scores from an instrument are stable and consistent (Creswell, 2012). The reliability of standardized reading achievement test scores is supported by the results of several studies of reliability conducted by ODE. An analysis of the standard error of measurement suggests that the
Test scores are reliable across the range of ability, except for the extreme ends of the distribution (ODE, 2007). No scores at the extreme ends of the distribution were detected in the data used in the study. The threat of violated assumptions of test statistics was controlled through testing the data to see if they meet the assumptions of the analysis. The low power threat was reduced by using a larger sample size (245 participants) and conducting power analysis. The results of the power analysis will be discussed in Chapter 4. Shadish et al. (2002) stipulated that “a poorly specified intervention point can seriously weaken the logic of the design” (p. 180). With respect to this study, this threat is controlled because the intervention point is clearly specified and the intervention was abrupt and permanent rather than gradual and/or temporary.

The threat to construct validity, according to Shadish et al. (2002), comes from the fact that, for the interrupted time series analysis, researchers may be forced to use outcome measures that are not fully relevant to the treatment being tested. In this study, the construct validity threat is controlled because the outcome measure (standardized reading achievement test) has been specifically designed to measure the dependent variable (reading achievement) in the study.

Bloom (2003) argued that when using the interrupted time series analysis it is important to control for cohort effects. Bloom (2003) described cohort effects as the factors that “influence the performance of students as a group rather than individually” (p. 10). Bloom (2003) stated that “little is known about cohort effects for short time-series of student test scores” (p. 13). However, according to Bloom (2003), it is possible to suggest that cohort effects may be larger for mathematics than for reading. This observation by Bloom (2003) and the fact that this study is focused on reading achievement contributed to minimizing the threat of cohort effects.

**Reliability of the research design.** Fraenkel et al. (2011) described reliability as the consistency of obtained scores. The reliability of the research design for this study is contingent
on the reliability of the instrument used for data collection. The instrument was the OAKS standardized reading achievement test. According to Fraenkel et al. (2011), reliability is determined by the degree to which scores consistently measure the variable the instrument is supposed to measure, which is reading achievement in the case of this study. The evidence for instrument reliability is derived from the analysis of the standard error of measurement (SEMeas). Fraenkel et al. (2011) described standard error of measurement as an index that estimates the degree to which a score would change under different circumstances. The average SEMeas for third grade reading tests was calculated at 3.2 RIT points on the basis of a normal distribution curve with a 90% confidence interval. The analysis of SEMeas suggested that the assessment provides similarly reliable test scores across the range of student ability with the exception of the extreme ends of the distribution (ODE, 2007). Accordingly, it is important to note that no scores at the extreme ends of distribution were observed in the data for the study.

**Validity of the language proficiency measure.** The details on the content, concurrent, and construct validity of the Woodcock-Muñoz Language Survey are provided in the *Woodcock-Muñoz Language Survey: Comprehensive Manual* (Woodcock & Muñoz-Sandoval, 1993a). The content validity of the language survey is supported by the fact that items included in various tests were selected on the basis of item validity studies and expert opinion. The purpose of each item is to adequately sample the ability measured by the test. Items of varying levels of difficulty were included in the survey to ensure that each test covers a wide range of ability. In addition, the cluster interpretation approach was utilized to improve content validity for broad abilities, such as oral language, reading, and writing and to minimize the possibility of drawing conclusions about language proficiency on the basis of a single, narrow aspect.

Woodcock and Muñoz-Sandoval (1993a) reported the results of several studies that
support the concurrent validity of the language survey. The reported studies ranged from a preschool study involving two- and three-year old children to a secondary school study involving students in grades 11 and 12. In the course of these studies, all subjects in the samples were administered other multiple batteries of tests in addition to the tests included in the survey. The concurrent validity correlations between the language survey and multiple test batteries demonstrated a wide range with a significant concentration of correlations in .70s, .80s and .90s.

Woodcock and Muñoz-Sandoval (1993a) stated that construct validity of a battery is demonstrated in how the measures of this battery are inter-correlated. Therefore, lower correlations signify dissimilarity of the constructs of different measures, and higher correlations signify similarity. Woodcock and Muñoz-Sandoval (1993a) argued that because all four tests measure different aspects of the broad construct of language proficiency, most tests are moderately inter-correlated at the level of .4 or above.

**Reliability of the language proficiency measure.** Internal consistency reliability coefficients and SEMeas were calculated to assess the reliability of the language survey. Woodcock and Muñoz-Sandoval (1993a) reported that “the test reliabilities were calculated by the split-half procedure, using odd and even raw scores, and corrected for length by the Spearman-Brown formula” (p. 53). In addition, cluster reliabilities were calculated using Mosier’s (1943) procedure. The reliabilities in the .80s and low .90s for individual tests and in the mid .90s for test clusters were established. The standard errors of measurement (SEMeas) varied from 5.4 to 6.2 for individual tests and were at the level of 4.0 for the oral language cluster, 4.2 for the reading-writing cluster, and 2.9 for the broad language proficiency.

**Generalizability.** Muijs (2011) noted that in quantitative studies researchers wish to generalize the findings from the sample to the population. Creswell (2012) stipulated that claims
about the generalizability of the study findings can be made based on the elimination of extraneous factors and the assumption that the outcome was the result of the treatment. Accordingly, caution should be exercised when attempting to generalize the results of this study. The need for this cautionary approach comes from the fact that the generalizability of the study results may be affected by the threats to external validity. Threats to external validity are related to the ability “to draw correct inferences from the sample data to other persons, settings, treatment variables, and measures” (Creswell, 2012, p. 306). Threats to external validity include interaction of selection and treatment and interaction of setting and treatment. Interaction of selection and treatment is related to the inability to generalize to other student populations. This threat was addressed by the fact that the treatment is relevant to a specific segment of the student population only, that is, elementary Spanish-speaking ELLs. The fact that generalization can only be considered with respect to this specific population is expected to reduce the possibility of error. Interaction of setting and treatment in this study is related to the inability to generalize to other school settings. As the school uses a specific reading instruction approach, a specific reading program, and specific screening procedures, it is important to exercise caution when considering generalization to other school settings leveraging different instructional approaches.

**Chapter Summary**

The study used the quantitative methodology in accordance with its goal to establish and explain the relationship between the variables. Quasi-experimental design was selected as the research design for the study. This design is appropriate for the study due to the lack of random assignment (Shadish et al., 2002). The study employed a convenience sampling method. The procedures required for the purpose of the study included obtaining relevant approvals for the use of the data and extracting the data from an existing database. The data consisted of the
student test scores on the third grade standardized English reading achievement test. Accordingly, the instrument for the study was the OAKS standardized reading achievement test. The data analysis plan included three steps. The first step was to prepare the data. The next step was to choose an appropriate statistical technique. As suggested by Shadish et al. (2002) and Tabachnick and Fidell (2007), short interrupted short time series method was used to estimate the causal effects of the change in the screening procedure on the later English reading achievement of ELLs. The final element of the data analysis plan was the analyses of the interrupted time series data. Given the nature of the research questions and the data, segmented regression was the statistical technique of choice that allowed the researcher to use interrupted time series analysis to assess the effects of the intervention on the outcome of interest.
CHAPTER 4: RESULTS

Research Questions

The purpose of this study was to determine whether placement of kindergarten ELLs either in L1 or English-only reading instruction based on their level of native language proficiency improves scores on a standardized reading achievement test in third grade when compared to the reading scores of third graders on the same test who, in previous years, were assigned either to L1 or English-only reading instruction without regard to L1 proficiency. To this effect, the study sought to answer the following research questions:

1. How, if at all, did the implementation of a new kindergarten screening procedure for Spanish-speaking ELLs regarding placement in L1 versus English-only reading instruction during the first four years of elementary school affect third grade student performance on the OAKS standardized reading achievement test?

2. How, if at all, did the implementation of a new kindergarten screening procedure for Spanish-speaking ELLs affect third grade student performance on the OAKS standardized reading achievement test for those placed in L1 versus English-only reading instruction?

Accordingly, this chapter presents the results of the statistical analyses that were conducted to determine the answers to each of the above research questions and to further explore the findings of the study. First, the chapter provides information about the preliminary analyses conducted with respect to the two research questions. These analyses include sample characteristics, power analysis, and descriptive statistics. Next, the findings of the inferential statistical analyses conducted to answer each of research questions are presented. Finally, the study results are summarized with respect to each research question.
Sample Characteristics

The initial available sample of ELLs who took the standardized reading achievement test in third grade during the period from 2008 to 2013 consisted of 332 students. After each ELL student who had a score on the third grade standardized reading achievement test was cross-referenced with the historical enrollment data, 87 students had to be excluded from the sample because they were enrolled at the school in grades other than kindergarten. Although these students took third grade standardized reading achievement test, later enrolment prevented them from being screened for English and, in later years, both for English and Spanish language proficiency to determine their placement either in L1 or English-only reading instruction. Upon enrollment, these students were placed in English-only reading instruction with the exception of nine students who, upon their enrollment at the school, were either monolingual Spanish-speakers or were enrolled in a bilingual reading program at their previous school. In addition to not having been screened for language proficiency to be considered for placement in L1 reading instruction, these students had only a partial exposure to the reading program used by the school. As for monolingual Spanish-speaking students, although they were enrolled in L1 reading instruction, they were not able to demonstrate an adequate level of English reading achievement on the standardized reading achievement test due to their limited English language proficiency.

The adjusted sample consisted of 245 students who were enrolled at the school in kindergarten, were screened to language proficiency upon entry, and continued to be enrolled at the school at the time of taking third grade standardized reading achievement test. All students were native Spanish speakers and qualified for ESL services. As about 80% of the student population at the school qualifies for free or reduced cost lunch and the ELLs account for 51% of the entire student population, it was safe to assume that most of 245 students in the sample come
from families with low socioeconomic status. The total adjusted sample included 116 boys (47%) and 130 girls (53%). The total number of students in three pre-intervention groups was 128 (52%), and the total number of students in three post-intervention groups was 117 (48%). The distribution of participants between the group that received L1 reading instruction and the group that received English-only reading instruction before and after the intervention was 98 (40%) and 147 (60%) students respectively. In terms of the distribution by gender, boys accounted for 50% (49 students) of the group that received L1 reading instruction and 46% (67 students) of the group that received English-only reading instruction. As can be seen from the above, the numbers in the sample were almost equally distributed in terms of gender and pre- and post-intervention attribution. The 40%/60% distribution between L1 and English-only reading instruction groups reflected the fact that the bilingual reading program was introduced at the school as a way to differentiate reading instruction for ELLs in order to support their English reading achievement while the majority of the ELLs would continue to receive English-only reading instruction. A detailed summary of the sample structure is provided in Table 4.
Table 4

Sample Structure Summary

<table>
<thead>
<tr>
<th>Student Groups</th>
<th>L1 Instruction</th>
<th>L2 Instruction</th>
<th>Total Students by Year</th>
<th>Total Students in Pre- and Post-intervention Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
<td>Total Students</td>
<td>Boys</td>
</tr>
<tr>
<td>Pre-intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(control)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>16</td>
<td>6</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>A2</td>
<td>16</td>
<td>7</td>
<td>9</td>
<td>34</td>
</tr>
<tr>
<td>A3</td>
<td>14</td>
<td>6</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>Post-intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(treatment)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>18</td>
<td>10</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>B2</td>
<td>21</td>
<td>13</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>B3</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Total Students by Gender</td>
<td>49</td>
<td>49</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>Totals Students by Language of Instruction</td>
<td>98</td>
<td>147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Students in Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Power Analysis

Creswell (2012) noted that sample size is one of the primary considerations that need to be taken into account when conducting a quantitative research study. Creswell (2012) argued that a larger sample will help to reduce a potential error when generalizing the study results from the available sample to the population. However, various factors may impose limits on the number of participants available to study. With respect to this study, a limited number of participants was conveniently available and met the sample selection criteria. Although in the
present study it was not feasible to obtain a larger sample, it was nevertheless important to find out whether the available sample was adequate for the study. An adequate sample size can be identified by conducting a power analysis. This procedure is aimed at determining a minimum sample size for making valid group comparisons on the basis of a certain level of statistical significance (alpha coefficient), effect size, amount of power, and number of variables.

The power analysis for this study was conducted on the basis of standard estimates. According to Creswell (2012), the alpha level is typically set at .01 or .05. An appropriate sample size for the study was computed on the basis of an alpha level of .05. Fraenkel et al. (2011) stipulated that a quantitative study will have a sufficient amount of power at a rigorous standard of .80. Effect size is defined by Fraenkel et al. (2011) as an index characterizing a degree of the magnitude of an obtained result of relationship (Fraenkel et al., 2011). According to Muijs (2011), $R^2$ is an effect measure used with respect to the multiple linear regression analysis. Based on the data from Cohen (1988), Kotrlik, Williams, and Jabor (2011) interpreted $R^2$ in the range from .1300 to .2600 as a medium effect size. The $R^2$ of .15 (medium effect size) was used to conduct a power analysis for this study. The power analysis was conducted for two numbers of variables in the model: 8-variable model and 9-variable model. As a result of the power analysis, the recommended sample size for an 8-variable model is 108, and for a 9-variable model it is 113. Consequently, it can be concluded that with an overall sample of 245 the study had adequate power even with the sample divided according to pre-intervention and post-intervention criterion and according to L1 and L2 reading instruction criterion.

**Descriptive Statistics**

The data analyses for this study were conducted using R. This program is a powerful
tool for performing statistical analyses and producing graphics (Teetor, 2011). The first step was to summarize the data using descriptive statistics. This statistics provides a meaningful description of the data using numerical indices and/or graphs (Fraenkel et al., 2011). In addition, descriptive statistics are used to determine overall trends and the distribution of the data (Creswell, 2012). The quantitative data in the study was represented by test scores on a standardized reading achievement test. This is a continuous dependent variable (Muijs, 2011). There are two types of categorical independent variables: (1) language of reading instruction, that is, Spanish (L1) and English (L2), and (2) group, that is, pre-intervention and post-intervention. Another independent variable is the year (1-6) when the OAKS reading test was taken. Table 5 shows a summary of the basic descriptive statistics for the continuous variable (test score). Table 6 shows frequency and relative percent for categorical variables (gender, language of instruction, year when the test was taken, and group).

Table 5

*Descriptive Statistics (Test score)*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>245</td>
<td>181.00</td>
<td>235.00</td>
<td>205.2776</td>
<td>8.90734</td>
<td>79.341</td>
</tr>
</tbody>
</table>
Table 6

*Frequency and Relative Percent (Categorical Variables)*

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>114</td>
<td>46.5</td>
</tr>
<tr>
<td>Female</td>
<td>131</td>
<td>53.5</td>
</tr>
<tr>
<td>L:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>98</td>
<td>40.0</td>
</tr>
<tr>
<td>L2</td>
<td>147</td>
<td>60.0</td>
</tr>
<tr>
<td>Year:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>38</td>
<td>15.5</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>20.4</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>16.3</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>18.4</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>16.7</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>12.7</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>128</td>
<td>52.2</td>
</tr>
<tr>
<td>Post</td>
<td>117</td>
<td>47.8</td>
</tr>
</tbody>
</table>

The segmented regression analysis that was used as an inferential statistics technique treats time as a continuous variable (Wagner et al., 2002). Therefore, year is regarded a continuous variable. Due to the fact that each participant belongs either to the control (pre-intervention) or treatment (post-intervention) group, the design of this study can be defined as between-groups (Creswell, 2012) or between-subjects design (Charness, Gneezy, & Kuhn, 2011). Table 6 is a summary of the between-subjects factors and the distribution of the sample among these factors.
In order to explore relationships within the data set it is necessary to use a measure of central tendency. A measure of central tendency is a summary number representing a single value in a distribution of scores (Vogt, 2005). According to Muijs (2011), the best measure of central tendency for a continuous variable, such as test scores, is the mean. In addition, the choice of the measure of central tendency is guided by the shape of the distribution (Manikandan, 2011). As it will be pointed out later in the chapter, the data distribution in this study roughly follows a normal curve. Consequently, the mean was selected as a measure of central tendency for this study. The estimated marginal means show the means of student test scores as a function of various student characteristics (e.g., gender, L1 or L2 reading instruction, reading test taken in years 1 through 6, and pre-intervention and post-intervention groups), as well as take into account various interactions among variables (Ho, 2006). The estimated marginal means are presented in Table 7. The range of marginal means is from 200.041 (in the second year pre-intervention year of the L1 group) to 212.384 (in the first pre-intervention of the L2 group) with an overall mean of 205.2776.
Table 7

*Estimated Marginal Means*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Males</td>
<td>204.563</td>
<td>.835</td>
<td>202.918</td>
<td>206.208</td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>205.309</td>
<td>.808</td>
<td>203.718</td>
<td>206.901</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>203.546</td>
<td>.896</td>
<td>201.780</td>
<td>205.312</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>206.326</td>
<td>.744</td>
<td>204.860</td>
<td>207.792</td>
<td></td>
</tr>
<tr>
<td>Pre-Intervention Group</td>
<td>205.152</td>
<td>.812</td>
<td>203.552</td>
<td>206.752</td>
<td></td>
</tr>
<tr>
<td>Post-Intervention Group</td>
<td>204.720</td>
<td>.830</td>
<td>203.084</td>
<td>206.356</td>
<td></td>
</tr>
<tr>
<td>Year 1 (pre-intervention)</td>
<td>208.227</td>
<td>1.441</td>
<td>205.387</td>
<td>211.067</td>
<td></td>
</tr>
<tr>
<td>Year 2 (pre-intervention)</td>
<td>203.922</td>
<td>1.331</td>
<td>201.300</td>
<td>206.544</td>
<td></td>
</tr>
<tr>
<td>Year 3 (pre-intervention)</td>
<td>203.307</td>
<td>1.450</td>
<td>200.449</td>
<td>206.165</td>
<td></td>
</tr>
<tr>
<td>Year 4 (post-intervention 1)</td>
<td>204.629</td>
<td>1.334</td>
<td>202.000</td>
<td>207.258</td>
<td></td>
</tr>
<tr>
<td>Year 5 (post-intervention 2)</td>
<td>204.978</td>
<td>1.367</td>
<td>202.285</td>
<td>207.671</td>
<td></td>
</tr>
<tr>
<td>Year 6 (post-intervention 3)</td>
<td>204.553</td>
<td>1.588</td>
<td>201.423</td>
<td>207.683</td>
<td></td>
</tr>
</tbody>
</table>
Inferential Statistics and Findings (RQ1)

The second step in the data analysis plan was to perform inferential statistical analyses with the idea of producing responses to the two research questions and testing the hypotheses derived from these questions. The ultimate objective of inferential statistical techniques is to determine how the findings from the sample used in the study can be generalized to the population (Fraenkel et al., 2011). The first research question was posed in order to find out if the implementation of a new kindergarten screening procedure for Spanish-speaking ELLs regarding placement in L1 versus English-only reading instruction during the first four years of elementary school affected third grade student performance on the OAKS standardized reading achievement test. The hypothesis derived from this research question posited that, following a change in the screening procedure for Spanish-speaking ELLs, third grade reading achievement significantly improved relative to that of students prior to the implementation of the new screening procedure.

Segmented regression analysis was conducted to answer the first research question. This analysis is a method within the framework of the multiple linear regression (MLR) using two or more independent variables. This technique helps to determine the relationship between an effect variable (also referred to as outcome or response variable), that is, a dependent variable and one or more predictors (also referred to as an explanatory variables), that is, independent variables (Muijs, 2011) in order to establish “how much an intervention changed an outcome of interest, immediately and over time” (Wagner et al., 2002, p. 299), or, in other words, to determine if the intervention effect was significantly greater than the underlying secular trend (Ramsay, Matowe, Grilli, Grimshaw, & Thomas, 2003). The secular trend is a tendency for the outcome to increase or decrease with time independently from the intervention (Ramsay et al., 2003). In an interrupted time series, a sequence of values of a
particular measure (student test scores on a standardized reading achievement test in this study) is interrupted or segmented. Segmentation occurs at a certain change point in time when the time series may demonstrate a change from the previously established pattern (Wagner et al., 2002). This segmentation may result from a real-world event, a change of policy, or an experiment (Wagner et al., 2002) that can all be regarded as interventions. With respect to this study, an intervention is represented by a change in the screening procedure that can be viewed as a change in policy. Wagner et al. (2002) stated that “segmented regression analysis is a method of statistically modelling the interrupted time series data to draw more formal conclusions about the impact of an intervention or event on the measure of interest” (p. 300).

Each segment in an interrupted time series is defined by the level and trend. The level is the value at the beginning of a time interval, that is, the $y$-intercept for the first or pre-intervention segment, and the value following a change point at which the segments join (Wagner et al., 2002). The trend, or the slope, is the rate of change in the measure of interest during the segment (Wagner et al., 2002). Wagner et al. (2002) stipulated that, in segmented regression analysis, each segment may exhibit both a level and a trend (slope). Therefore, researchers need to examine changes in the level and/or trend by using statistical models estimating the level and trend in the pre- and the post-intervention segments. As per Wagner et al. (2002), the specification of segmented regression is as follows:

$$Y_t = \beta_0 + \beta_1 \times \text{time}_{t} + \beta_2 \times \text{intervention} + \beta_3 \times \text{time after intervention}_{t} + \epsilon_{t}$$

In this formula, $Y_t$ is the outcome variable at time $t$. Time is a continuous variable from the beginning to the end of the study. Wagner et al. (2002) suggested that time can be coded as 0 for the first observation before the intervention. Therefore, $\beta_0$ is an intercept and provides an estimate for the baseline level of the outcome variable at time 0, whereas $\beta_1$ is an estimate of the change in
the outcome of interest that occurs during pre-intervention observation. Accordingly, this is the baseline trend. Further, $\beta_2$ estimates the level of change immediately after the intervention from the end of the pre-intervention segment, and $\beta_3$ estimates the change in the trend after the intervention, compared to the trend before the intervention. The sum of $\beta_1$ and $\beta_3$ is the post-intervention slope. Wagner et al. (2002) argued that the use of the above model gives researchers an opportunity to control for the baseline level and trend. This control is “a major strength of segmented regression analysis” (Wagner et al., 2002, p. 301). Finally, $\epsilon_t$ at time $t$ indicates the random variability not explained by the model.

The above model was applied to the study with the use of the codes explained below. In this study, the dependent variable is a continuous variable represented by student score on the standardized reading achievement test with the English reading achievement as the outcome of interest. The independent variables or predictors are as follows: (1) L- language of instruction, (2) Group-pre- and post-intervention, (3) Year (i.e., annual test administration during six consecutive years).

To answer the first research question, segmented regression analysis was conducted to model the variation in scores before and after the intervention with respect to all ELLs. Although in this model the means of scores for L1 and L2 are examined together, the model contains a variable controlling for the group differences. To this effect, R was used to fit a segmented regression model with the test score as the response variable and time, group (pre- and post- intervention), and language of instruction as the explanatory variables. The output obtained on the basis of the available sample is shown in Table 8.
Table 8

Results of Segmented Regression Analysis (Overall Model)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>(t) value</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>209.5862</td>
<td>94.043</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>L</td>
<td>2.9571</td>
<td>2.592</td>
<td>0.01013</td>
</tr>
<tr>
<td>Year</td>
<td>-2.7927</td>
<td>-2.830</td>
<td>0.00505</td>
</tr>
<tr>
<td>Group</td>
<td>1.7554</td>
<td>0.729</td>
<td>0.46677</td>
</tr>
<tr>
<td>Year After Intervention</td>
<td>2.8157</td>
<td>1.993</td>
<td>0.04740</td>
</tr>
</tbody>
</table>

Note. \(L\) reference = \(L1\); Group Reference = Pre-Intervention
\(R^2 = .06;\) Adj. \(R^2 = .04434; F(4, 240) = 3.83, p = .004884\)

The first step in performing statistical inference using the output of segmented regression was to check if the explanatory variables (time, language, and group) had an effect on the response variable (test score), that is, test the null hypothesis positing that the model coefficients are equal to zero (Warner, 2007). This can be done by examining \(p\)-value established for each coefficient. Each coefficient with \(p < .05\) corresponds to a variable that explains a significant portion of unique variance in the outcome variable. The results of an omnibus test warranted the rejection of the null hypothesis positing that these predictor variables in combination were not explaining any variation in model \(F(4, 240) = 3.83, p < .01\). Thus, it was necessary to evaluate each coefficient separately in order to determine each variable’s significance in the model. The analysis demonstrated that the language of instruction \((t = 2.592, p = 0.01013)\), year \((t = -2.830, p = 0.00505)\), and year after intervention \((t = 1.933, p = 0.04740)\) were significant predictors of the score outcome. To the contrary, the results of the group predictor proved to be non-significant \((t = 0.729, p = 0.4667)\) and, as a result, this predictor cannot be viewed as significant with respect to the mean of test score in terms of the overall sample. From the formula, it can be seen that the
change in the mean score between the group that received L1 reading instruction and the group that received L2 reading instruction was 2.9571. This number indicates the mean difference in scores between the groups that received either L1 or L2 reading instruction, with the L2 group showing higher scores than the L1 group. The second part of the formula demonstrates the trend before the intervention. The equation -2.7927*Year reflects a negative trend in scores during the pre-intervention period with the mean score decreasing by an average of 2.7927 each year. The negative trend was reversed following the intervention, as the change in the mean of scores following the intervention was positive and sustained (2.8157*Year_After_Intervention-2.7927*Year). Figure 7 provides a graphic illustration of the model using estimated marginal means of score highlighting that the intervention occurred after the third year.

![Figure 7. Mean Scores Through Time (Overall Model).](image)

The next step was to preform model diagnostics. First, it was necessary to determine the spread, that is, the amount of variability in the distribution of the data (Creswell, 2012).
Many distributions of the data have a tendency to be normal, that is, follow a normal curve that is bell-shaped (Fraenkel et al, 2011). Fraenkel et al. (2011) stipulated the importance of normal distribution for many inferential statistical techniques. This importance is also suggested by Frees (2009) who argued that normal distribution of data is “pivotal” for conducting regression analysis used as an inferential statistical technique in this study. According to Frees (2009), quantile-quantile or q-q plots can help researchers assess the normality of a sample using the normal curve as a reference distribution. In this case, the vertical axis demonstrates quantiles for each point of the sample distribution, and the horizontal axis demonstrates the corresponding quantity based on the normal curve. If the quantile points of the sample curve lie along the line connecting the 25th and the 75th points of the normal distribution curve, the sample and normal distribution have the same shape (Frees, 2009). The q-q plot for the overall model shown in Figure 8 provides a visual illustration of the spread and the data normality.

However, while graphical methods, such as quantile-quantile plots are the easiest and commonly used diagnostic tool for checking the data normality, “they are still not sufficient to provide conclusive evidence that the normal assumption holds” (Razali & Wah, 2011, p. 21). Therefore, graphical methods of ascertaining the normality of data need to be supported by formal numerical methods. Numerical methods imply the use of formal normality tests that should be performed prior to making assumptions about data normality (Razali & Wah, 2011). As a formal normality test, this study used the Shapiro-Wilk (SW) test. First, the SW test was performed with respect to the normality of the model for the overall sample. Based on the Shapiro-Wilk test statistic ($W = .9896$) and $p$-value that is more than .05 ($p = .07676$), the conclusion was that the overall sample comes from a normal distribution.
In order to have reliable results when analyzing data it is also important to confirm that various other assumptions of the segmented regression are aligned with the data used in the study. This can be done via the GVLMA package in R. The GVLMA enables a comprehensive check of the linear model, including the fit, the shape of the distribution of the residuals (skewness and kurtosis), as well as the linearity and heteroscedasticity (Peña & Slate, 2008). The GVLMA output for the overall sample demonstrated that the linear model did not meet the assumptions in terms of two out of five coefficients. However the residuals of the linear model in this case met the assumptions in terms of skewness and kurtosis, in addition to heteroscedasticity. Given the appropriateness supported by q-q plots, normality test, skewness, kurtosis, and heteroscedasticity diagnostics, it was possible to conclude that any violations of model assumptions, if present in this model, can be regarded as mild and do not provide any
cause for concern. As such, the use of this model is supported by the underlying statistical theory (Box & Watson, 1962).

In view of the foregoing, it can be concluded that a change in the screening procedure used for placement ELLs either in L1 or L2 reading instruction appeared to have a moderately positive effect on the reading achievement of all ELLs in the sample. This effect is demonstrated by the fact that although a sharply negative trend in the means of test scores was observed in the pre-intervention group, the negative pre-intervention trend was stabilized after the intervention (change in the screening procedure). However, while the post-intervention trend is positive, the slope is nearly zero. Consequently, the research findings in part supported the hypothesis derived from the first research question which posited that a change in the screening procedure will lead to a significant improvement of reading test scores. Although there was a statistically significant relationship between the response and explanatory variables as demonstrated by the difference in trend between the pre- and post-intervention periods, no statistically significant difference in the means of test scores was observed.

**Inferential Statistics and Findings (RQ2)**

The goal of the second research question was to find out if the implementation of a new kindergarten screening procedure used for placing Spanish-speaking ELLs either in L1 or L2 reading instruction during the first four years of elementary school differently affected the performance of students in the L1 group and the performance of students in the L2 group on the OAKS standardized reading achievement test in third grade. The hypothesis derived from this research question posited that, following a change in the screening procedure, third grade reading achievement of ELLs who received L1 reading instruction significantly improved relative to that of ELLs who received L2 reading instruction. Segmented regression analysis with the use of R
was conducted to answer the second research question as well. A total of three models were created to explore this hypothesis. The first model was a segmented regression analysis built exclusively on the L1 group; the second model was a segmented regression analysis built exclusively on the L2 group; and, the third model was a modification of the segmented regression model built for the first research question to include an interaction between time and the language of instruction variable (L1/L2). The output shown in Table 9 reflects estimates built exclusively on the L1 model obtained on the basis of the available sample.

Table 9

Results of Segmented Regression Analysis (L1 Model)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>203.8241</td>
<td>67.635</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Year</td>
<td>-0.8212</td>
<td>-0.577</td>
<td>0.566</td>
</tr>
<tr>
<td>Group</td>
<td>3.1159</td>
<td>0.903</td>
<td>0.369</td>
</tr>
<tr>
<td>Year After Intervention</td>
<td>0.9346</td>
<td>0.466</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Note. Group Reference = Pre-Intervention

$R^2 = .02896$; Adj. $R^2 = -.002$; $F(3, 94) = 0.9345, p = .4273$

The omnibus test suggested that neither time nor group as explanatory variables were statistically significant with respect to the mean of the test scores. Figure 9 provides a graphic illustration of the model using estimated marginal means of score.
The next step was to analyze the output presented in Table 10. These results pertain to ELLs who received L2 reading instruction.

Table 10

*Results of Segmented Regression Analysis (L2 Model)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>216.4857</td>
<td>75.499</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Year</td>
<td>-4.1954</td>
<td>-3.199</td>
<td>0.0017</td>
</tr>
<tr>
<td>Group</td>
<td>0.9483</td>
<td>0.295</td>
<td>0.7685</td>
</tr>
<tr>
<td>Year After Intervention</td>
<td>4.0706</td>
<td>2.146</td>
<td>0.0336</td>
</tr>
</tbody>
</table>

*Note.* Group Reference = Pre-Intervention  
$R^2 = .09478$; Adj. $R^2 = .07579$; $F(3, 143) = 4.991, p = .002538$
The omnibus test suggested an existence of a significant relationship between the response and at least one of the explanatory variables, that is, the mean of test score for the L2 group as the response variable and time and group (pre- and post-intervention) as explanatory variables \( F(3, 143) = 4.991, p = .0025 \). The pre-intervention trend for the L2 group was negative \((-4.1954\times \text{Year})\). The post-intervention trend for the L2 group showed a significant improvement while at the same time remaining negative \((4.1954+4.0706 = -1.248)\). A graphic illustration of the trend for the L2 group is shown in Figure 10.

![Figure 10](image)

*Figure 10.* Mean Scores Through Time (L2 Model).

In view of the above, it can be stated that the change in the screening procedure did not lead to a statistically significant difference either in trend \((p = .566; \ p = .642)\) or mean test scores \((p = .369)\) for those ELLs who received L1 reading instruction, as reflected in the L1 model. At the same time, a statistically significant relationship was observed between two of the explanatory variables and the response variable in the L2 model. However, the inferential
statistics results demonstrate that this statistically significant relationship was limited to the slopes before (Year) \( p = .0017 \) and after (Year _ After _ Intervention) \( p = .0336 \) the intervention. This indicates that the intervention had an effect on the change of slope between the pre- and post-intervention periods, but the raw effect of Group \( p = .7685 \), the variable measuring the mean difference between pre- and post-intervention periods, was not statistically significant. In other words, a statistically significant relationship between time and the test scores of the L2 group was observed in the form of different slopes for the pre- and post-intervention periods. This trend, negative in the pre-intervention years, improved dramatically, though remaining negative, in the years following the intervention.

Although no statistically significant difference in the means between the pre- and post-intervention groups either in the L1 or L2 models was observed, the early changes in trends of the estimated marginal means of test scores of the L1 and L2 pre- and post-intervention groups suggest there may be a statistically significant difference in how the two groups responded to the intervention. To test this theory and, thereby address the second research question, it was necessary to use a model with an interaction term component.

In regression models, interaction terms are products of two or more predictors (Gunst & Mason, 1980). Gunst and Mason (1980) argued that interaction terms should be introduced when there may be reason to believe that the effect of a predictor on the outcome is moderated by the values of other predictors. In such cases, according to Gunst and Mason (1980), interaction terms provide more opportunities for individual predictor variables to demonstrate joint effect with other predictor variables. With respect to this study using the data collected over a period of six year, it was appropriate to hypothesize that the effect of one predictor (time after intervention) on the outcome (OAKS test scores) was moderated by the values of another
predictor (L1/L2). In other words, an interaction term would provide an opportunity to effectively measure whether the means and trends of test scores for the L1 and L2 groups reacted differently to the intervention. In terms of the time series analysis, an interaction term allows for an estimate of whether the trends for the L1 and L2 groups statistically diverged (or converged) after the intervention. That is, an interaction terms would permit the researcher to formally look for differences between the post-intervention trends for the L1 and L2 groups. This formal statistical comparison of these trends was a way to generate an answer to the second research question and test the relevant hypothesis. The output for the interaction model is shown in Table 11.

Table 11

*Segmented Regression Output (Interaction Model)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>208.2904</td>
<td>91.537</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>L</td>
<td>5.2291</td>
<td>3.518</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Year</td>
<td>-2.8719</td>
<td>-2.936</td>
<td>0.004</td>
</tr>
<tr>
<td>Group</td>
<td>1.9195</td>
<td>0.804</td>
<td>0.422</td>
</tr>
<tr>
<td>Year After Intervention</td>
<td>4.2882</td>
<td>2.797</td>
<td>0.0056</td>
</tr>
<tr>
<td>Year After Intervention *L</td>
<td>-2.4298</td>
<td>-2.353</td>
<td>0.0194</td>
</tr>
</tbody>
</table>

*Note.* L reference = L1 Group Reference = Pre-Intervention

\( R^2 = .08129; \text{ Adj. } R^2 = .06207; F(5, 239) = 4.23, p = .001 \)

A review of the above model showed that, with the exception of one variable, all explanatory variables exhibited a statistically significant relationship with the response variable \((p < .05)\). The only non-significant explanatory variable was the group variable \((p = .422077)\).
Thus, the interaction model is as follows:

\[
\text{Score} = 208.2904 - 2.8719 \times \text{Year} + 1.915 \times \text{Group} + 4.2882 \times \text{Year}_\text{After}\_\text{Intervention} \\
+ 5.2291 \times \text{L1} - 2.4298 \times \text{Year}_\text{After}\_\text{Intervention} \times \text{L1}
\] (2)

A review of the equation revealed a statistically significant difference in the change of pre- and post-intervention trends between the L1 and L2 groups: \(-2.4298 \times \text{Year}_\text{After}\_\text{Intervention} \times \text{L} \ (p = .0194)\).

The difference in trends is demonstrated by the fact that a change in slope of the L2 group was 2.4298 points less than that of the L1 group for the years following the intervention. This difference in trends between L1 and L2 groups warranted a conclusion that the effect of the intervention was statistically different between the two groups. Accordingly, it is possible to say that, as a result of the intervention, the reading achievement of the L1 group significantly improved relative to that of the L2 as suggested by the annual trends of the mean test scores of each group. However, as no statistical difference was observed in the mean test scores of pre- and post-intervention groups both with respect to the L1 and L2 models, it should be stated that the second research hypothesis was supported only partially.

To complete the review of the L1 and L2 models as well as the interaction model that was used to formally compare L1 and L2 groups over time, it was necessary to perform model diagnostics for these three models. The first element of the model diagnostics was the data normality check. Figures 11, 12, and 13 provide a visual illustration of the spread and the data normality by showing quantile-quantile plots for the three models.
Figure 11. Q-Q Plot (L1 Model).

Figure 12. Q-Q Plot (L2 Model).
Figure 13. Q-Q Plot (Interaction Model).

In addition, the SW tests were performed with respect to the normality of the three models. The output for the tests is shown in Table 12.

Table 12

Results of Shapiro-Wilk Normality Test.

<table>
<thead>
<tr>
<th>Model</th>
<th>Shapiro-Wilk Test Statistic $(W)$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>.982</td>
<td>.2019</td>
</tr>
<tr>
<td>L2</td>
<td>.9845</td>
<td>.09861</td>
</tr>
<tr>
<td>Interaction</td>
<td>.9875</td>
<td>.0313</td>
</tr>
</tbody>
</table>

The test results demonstrated $p = .2019$ for the sample of ELLs who received L1 reading instruction and $p = .09861$ and for the sample of ELLs who received L2 reading instruction.
Given that these \( p \)-values are more than .05, the conclusion was that these samples came from a normal distribution. Although the results for the interactions model showed \( p \)-value = .0313 signifying that these data may not have come from a normal distribution, it can be shown that linear models are relatively robust to small departures from normality (Box & Watson, 1962). Given the borderline \( p \)-value and the quality of the quantile-quantile plot, it can be concluded that this departure from normality is indeed small and, thus, not a threat to the validity of the inference made from this model.

Next, an assessment of the model fit was required. In order to meet this requirement, residual diagnostics via the GVLMA package in R were performed. With respect to the L1 and L2 models, all assumptions related to linear regression diagnostics were acceptable (\( p > .05 \)) suggesting that the inference from these models is valid.

To confirm the results of the interaction model used to formally compare L1 and L2 groups over time, an assessment of the model fit was also required. The results of the diagnostics for the interaction model indicated that the model fit the data with respect to two coefficients (skewness and heteroscedasticity). The model did not demonstrate fit with the data in terms of global stat, kurtosis, and link function. Nevertheless, based on the appropriateness supported by q-q plots, normality tests, skewness, and heteroscedasticity diagnostics, it was safe to conclude that any violations of model assumptions, if in fact present in the model, can be viewed as mild and do not create any cause for concern. As such, it can be concluded that the use of the model is supported by the underlying statistical theory (Box & Watson, 1962).

**Chapter Summary**

This chapter presented the results of data analyses conducted within the framework of the study. First, the criteria used for including the participants in the sample were outlined, and the
sample characteristics were identified. Based on the sample selection criteria, 245 ELLs were included in the sample. The overall sample included four independent variables: group (pre- and post-), time (years 1 through 6), post-intervention years (years 4 through 6), and language of instruction (L1 or L2). Next, the power analysis was performed to ensure that the sample sizes were adequate. The results of the power analysis confirmed that there was sufficient power in terms of the overall sample and the sample subsets.

Once the sample was fully identified and calculated and the results of the power analysis confirmed adequate sample power, it was possible to apply descriptive and inferential statistics techniques to the data that were extracted from the OAKS database and coded. In terms of descriptive statistics it was essential to define a measure of central tendency and determine the distribution of the data. In addition, it was necessary to check if the sample came from the population with a normal distribution. Four Shapiro-Wilk tests were performed for this purpose. The results of the SW test confirmed the normality of data distribution for three out of four models. The fourth model, that is, the interaction model, that failed the SW normality test, showed only mild departures from the normal distribution, and its associated quantile-quantile plot was satisfactory. The foregoing made it possible to assume approximate normality that posed no serious threat to the model’s validity. Quantile-quantile plots generated for the overall model, the L1 and L2 models, and the interaction model provided a graphic illustration of the normal distribution.

After summarizing the data via descriptive statistics it was appropriate to conduct inferential statistical analyses in order to generate responses to the research questions stipulated in the study and to determine if the results of the inferential statistical analyses would support the hypotheses originating from the research questions. To this effect, it was necessary to start with
identifying an inferential statistical method that would be the best possible fit with the data. It was determined that the segmented regression analysis, a method existing within the framework of the multiple linear regression, would be an appropriate option. The segmented regression analyses were performed by an outside statistical consultant who conducted the tests with the use of the R software package. The output of the segmented regression models was analyzed with respect to the research questions. Additional model diagnostics were conducted to determine if the selected models fit with the data used in the study. Out of the four models, two models demonstrated acceptable assumptions on all five coefficients, and the other two models demonstrated acceptable assumptions on three and two out of five coefficients. However, the deviations were mild and given the satisfactory q-q plots, heteroscedasticity, skewness, kurtosis, and normality diagnostics, these slight deviations should not be considered as a threat to model validity. Therefore, the results suggested that the overall model performance was satisfactory.

Analysis of the first segmented regression model, that is, the model with no interaction effect, led to the conclusion that the hypothesis associated with the first research question was only partially supported. Namely, it was supported in terms of a stabilized and positively altered trend, but not in terms of the mean test scores. The hypothesis related to the second research question was also partially supported as the interaction model confirmed a statistically significant difference between the trends of the L1 and L2 groups after the intervention, but no statistically significant difference in the mean test scores of the L1 and L2 groups was discovered. In particular, the model showed that the change in slope for the L1 group was larger than the change in slope for the L2 group. The next chapter will discuss how the results of the study pertain to the theory and practice related to the reading instruction of ELLs and their English reading achievement.
CHAPTER 5: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter is the concluding chapter where the results of the current study are summarized and relevant conclusions are drawn from the results. In order to provide context for the results and conclusions the chapter begins with a restatement of the research problem underlying the study and a brief outline of the theoretical rationale developed for the study. In addition, while drawing conclusions based on the study results it is important to take into account the limitations of the study. The conclusions drawn from the study findings provide a foundation for discussion as well as determine theoretical and practical implications of the study. In turn, theoretical and practical implications help to provide recommendations for future research. Final comments complete the chapter.

Restatement of the Problem

The purpose of this study was to investigate to what extent a change of the screening methodology used to place incoming kindergarten students identified as ELLs either in Spanish (L1) or English (L2) reading instruction affected their English reading achievement as measured by the scores on a standardized English reading achievement test taken in third grade. The study aimed at examining the effects of the change in the screening methodology from two perspectives: (1) with respect to all ELLS regardless of the language in which they received reading instruction; (2) with respect to each separate group of ELLs, that is, ELLs who received L1 reading instruction and ELLs who received L2 reading instruction. Finally, the study looked at the differences in the effect of the change in the screening procedures between the group of ELLs that received L1 instruction and the group that received L2 instruction.

The purpose of the study was driven by several well-documented factors. First, over recent decades, the population of ELLs has been on a continuous increase. This statement is
supported by considerable data from a variety of sources (e.g., National Clearing House for English Language Acquisition, 2006; Office of English Language Acquisition, Academic Enhancement, and Academic Achievement for Limited English Proficiency Students, 2008). Based on the data provided by the U.S. Department of Education (U.S. Department of Education, Office of the Secretary, 1991), it is possible to say that the ELL population in U.S. schools has more than doubled during the last 25 years (García, Kleifgen, & Falchi, 2008). ELLs living in the United States speak more than 460 languages (Kindler, 2002). However, among this variety of languages Spanish is clearly the most common one. According to various estimates, Spanish is the primary language for 75% to 79% of the overall population of ELLs (Kindler, 2002; Zehler et al., 2003). Another well-documented and disturbing fact is the academic underachievement of Spanish-speaking ELLs. Figueras-Daniel and Barnett (2013) observed that the existing statistics on the academic performance of Spanish-speaking ELLs cannot be viewed as encouraging. Indeed, according to the 2011 data from the National Center for Education Statistics, only 14% of Spanish-speaking ELLs demonstrated adequate reading proficiency in fourth grade. In view of this situation, English reading of ELLs comes to the forefront of issues that require the attention of educators, researchers, and policy makers.

Within the scope of the general question of how to promote English reading achievement of all ELLs and Spanish-speaking ELLs in particular, the issue of the language of reading instruction has taken a central position. Mathes et al. (2007) claimed that most of the research with respect to the instruction of ELLs was focused on the language of instruction. The importance of the issue of the language of reading instruction was explained by Bernhardt (2003) who argued that it was important for the literacy community to understand “the complexities of learning to read in a language when another literacy or another oral language exists in cognition”
In the 2009 review of the NRP Report (August & Shanahan, 2006), Escamilla (2009) emphasized the fact that the report clearly confirmed a higher effectiveness of bilingual approaches to reading instruction of ELLs compared to English-only approaches. This conclusion was supported by the results of other research studies (e.g., Slavin & Cheung, 2005; Rolstad et al., 2005; Branum-Martin et al., 2010).

The existing body of research on L1 reading instruction of ELLs and specifically Spanish-speaking ELLs has been abundant and multi-faceted. Accordingly, several major themes can be identified in the research literature. The first one is related to the general issue of L2 literacy development as opposed to L1 literacy development. Quite a few authors (e.g., Bernhardt, 2003; Escamilla, 2009; Hammer et al., 2011) emphasized the importance of not equating English language literacy development of ELLs to that of native English speakers and argued that L2 literacy should not be viewed exclusively from a monolingual perspective, therefore, calling for additional research into the instructional methods that would be specifically applicable to ELLs.

One method that stands out in this respect is the use of students’ L1 as the language of initial reading instruction, that is, various types of bilingual reading programs. There have been multiple research studies with the goal of determining the efficacy of bilingual reading programs (e.g., Rolla San Francisco et al., 2006; Nakamoto et al., 2010; Branum-Martin et al., 2010) as well as meta-analyses evaluating such research studies (e.g., Slavin & Cheung, 2005; Rolstad et al., 2005). Based on the review of this category of research literature, it is safe to say that over the recent years the majority of researchers were in agreement about the fact that in terms of literacy instruction of ELLs “bilingual approaches are more effective than English only approaches” (Escamilla, 2009, p. 451).
The third major theme detected in the literature on L1 literacy instruction of ELLs is related to the CLT of literacy skills from L1 to L2. This type of research is based on the concept of CLT supported by the premise of linguistic interdependence (Cummins, 1979). The concept of CLT provides theoretical foundation for bilingual education in general and for L1 reading instruction in particular. There is an extensive body of research on CLT with respect to various reading skills (e.g., phonemic awareness, decoding, and reading comprehension). Research findings on the CLT of reading skills ranged from no evidence of CLT (e.g., Mathes et al., 2007) to a modest level of CLT (e.g., Manis et al., 2004), to comprehensive and sustained CLT (Sparks et al., 2008). It was also observed that the extent of CLT may vary for specific reading skills as well as be positively correlated with the level of L1 proficiency.

Although there is an abundance of research studies on the reading instruction of ELLs, there is a clear need for additional more targeted research. While the existing research appears to have provided sufficient evidence for the overall beneficial role of initial L1 reading instruction for later L2 reading achievement of ELLs, more research findings are needed to provide a comprehensive understanding of various factors that may either enhance or impede CLT of reading skills from L1 to L2. The NRP Report (August & Shanahan, 2006) noted that individual differences among ELLs could make a significant contribution to their English literacy development. Given the complexity of the literacy development process with respect to ELLs, there are many individual factors that can influence the dynamic process of their English literacy development (August & Shanahan, 2006).

The levels of language proficiency both in L1 and L2 are among such factors. Durgunoğlu (2002) suggested that further research was needed to determine if and how initial L1 proficiency may affect CLT from L1 to L2. This statement is supported by the threshold
hypothesis by Cummins (1979) alleging that a certain level of linguistic skills in L1 may be needed to support linguistic interdependence and CLT. Mathes et al. (2006) reasoned that it would be naïve to assume that information learned in one language will automatically translate into knowledge in another language. Therefore, Mathes et al. (2006) noted the need for more research looking into what would be necessary to ensure such transfer. Based on the assumption by Durgunoğlu (2002), the threshold hypothesis by Cummins (1979), and an observation by Mathes et al. (2006), it would be logical to conclude that one factor that should be carefully examined with respect to supporting English literacy development of ELLs via initial L1 reading instruction is the initial level of L1 proficiency.

However, to date, there appears to be no research studies specifically addressing this issue. Cárdenas-Hagan et al. (2007) contended that initial reading instruction for ELLs should take into account both L1 and L2 skills and decisions about the language of instruction should be made with the goal of enhancing student learning. Nonetheless, there are no studies looking into how decisions about the language of reading instruction for ELLs may be made to maximize the student learning potential. While L1 reading instruction may be beneficial for ELLs, it cannot be expected that “a one-size-fits-all approach will create able readers” (Atwill et al., 2011). Atwill et al. (2011) maintained that one should not assume that all young ELLs have identical linguistic experiences and backgrounds. Therefore, it would merit finding out more about students’ linguistic status when they enter school, so that teachers are able to make more informed and individualized decisions about what language of reading instruction may be more appropriate for each student. This study attempted to fill in the gap in the existing body of research with respect to the foregoing matter.

To summarize, the purpose of the present research was to investigate the effects of the
change in the screening methodology used for placing kindergarten Spanish-speaking ELLs either in L1 or L2 reading instruction. A review of the existing literature revealed that, although there have been multiple research studies and meta-analyses on the subject of the efficacy of L1 reading instruction for ELLs as a way to enhance their English language literacy achievement as well as multiple studies on the CLT of various reading skills from L1 to L2, to date, there has been no research that directly addressed the issue of placing kindergarten ELLs either in L1 or L2 depending upon their L1 proficiency.

**Restatement of the Theoretical Rationale**

The theoretical rationale for this study is based on a synthesis of the seminal works by Cummins, a leading theorist in and a consistent advocate of bilingual education. Several constructs developed by Cummins are directly congruent with this research study. The first is the CUP model (Cummins & Swain, 1986) that provides a foundation for the understanding of bilingual proficiency. According to the CUP model, L1 and L2 proficiencies are not isolated from one another, but, on the contrary, share a common underlying proficiency. Although this common underlying proficiency may not be immediately evident and L1 and L2 may appear to be independent from each other, it provides basis for linguistic interdependence between L1 and L2 because, if L1 and L2 share a common underlying proficiency, it is logical to conclude that the two language proficiencies interact with each other. The principle of linguistic interdependence is the next theoretical construct by Cummins (1981).

This principle posits that knowledge and skills acquired in L1 can have an effect on knowledge and skills in L2. Thereby, the principle of linguistic interdependence is directly related to the construct of CLT that can be regarded as fundamental with respect to second language learning and bilingual education. The principle of linguistic interdependence
conceptualizes the construct of CLT at the level of skills that are more cognitive in nature. In addition, Cummins (1980b) developed the constructs of BICS and CALP that represent different types of linguistic proficiency: linguistic skills needed for everyday communicative language use (BICS) and more cognitively demanding linguistic skills necessary to perform academic tasks (CALP). These constructs of the common underlying proficiency, linguistic interdependence, and CALP, combined with the concept of CLT, create a theoretical framework in support of bilingual education including L1 reading instruction.

When considering the option of L1 instruction it is also important to take into account the threshold hypothesis by Cummins (1979). This hypothesis alleges that the extent of CLT, particularly of CALP-related knowledge and skills, is influenced by the level of linguistic proficiency in L1. In other words, there is a certain “threshold” level of L1 language skills that needs to be achieved in order to make CLT to L2 possible. All the foregoing constructs are closely related and in combination provide a theoretical rationale for bilingual education. Therefore, they can be regarded as an appropriate choice of a theoretical rationale for this study.

**Limitations**

Prior to drawing conclusions about the study findings, it is necessary to define limitations that may have affected the results (Creswell, 2012). The reason for this is the fact that an explanation of limitations is instrumental to an understanding of the extent to which the study findings can or cannot be generalized to different populations and settings.

The first limitation is the sample size. Creswell (2012) argued that it is always advisable to select a sample size as large as possible because a larger sample will help to minimize the potential error of the sample being different from the population. However, Creswell (2012) admitted that various factors can impose limitations on the sample size. In relation to this study,
availability and the number of variables (Creswell, 2012) constitute such factors. The fact that this study aimed to examine the effects of a specific intervention related to the change in the screening methodology used to place kindergarten ELLs either in L1 or L2 reading instruction automatically imposed a constraint in terms of the availability of participants. Multiple independent variables (language of instruction, pre/post intervention, and year) also acted as limiting factors. Moreover, to ensure the validity of the study results, the sample had to be reduced from 332 to 245 ELLs because 87 potential participants did not meet a critical inclusion criterion (initial language proficiency screening upon entering kindergarten) due to their later enrollment at the school. Although the power analysis conducted for two numbers of variables in the model (an 8-variable model and a 9-variable model) confirmed that the sample size was adequate for both models, a larger sample would have helped to further minimize sampling error and allow for more robust and accurate generalizations from the study results.

The second limitation pertains to the number of observations in the interrupted time series data used in the study. As it was mentioned in Chapter 3, Box et al. (1994) originally insisted on a large number of observations to identify a stable model in the interrupted time series analysis. Later, Shadish et al. (2002) modified this requirement by introducing the concept of the abbreviated time series. Wong et al. (2009) expanded on the work of Shadish et al. (2002) by suggesting that the abbreviated interrupted time series (AITS) may be used in the education setting. Wong et al. (2009) contended that, in the education setting, a large number of pre- and post-intervention observations could be highly problematic. With respect to the interrupted time series data in this study, the need to use the same instrument for data collection limited the number of available observations. In other words, there was a specific window of opportunity lasting for six years (from 2007/2008 to 2012/13 school years) when ODE consistently used the
same standardized reading achievement test (OAKS test). While the NLL program at the school continued to operate, additional observations were not possible due to the change of the standardized reading test by ODE. In summary, although the study used the data from a number of observations that is regarded as acceptable with respect to the AITS, a larger number of observations would be desirable to observe changes in the trends over time.

Although every effort has been made to account for any confounding factors, such as participant profiles, reading instruction delivery over time in terms of instructional approaches and teachers, it was not possible to account for how individual teaching styles may have positively or negatively influenced English reading achievement outcomes over the nine-year time span of the program. In addition, while it is known that most ELLs are not likely to receive academic support at home, over the years there may have been a number of ELLs who were supported in their literacy learning at home either in L1 or L2, or both. Within the framework of the study, it was not possible to control for this factor. Finally, even with the homogeneity of the sample, it was not possible to determine how individual student differences may have influenced their English reading achievement over time. Given the longitudinal aspect of the study, the foregoing variations (e.g., individual teaching styles, extent of family support, student individual profiles) may have had a certain cumulative effect that could not be quantified, but clearly needs to be acknowledged. This consideration is the third limitation in the study.

The fourth limitation is related to making generalized causal inferences regarding the relationships among the study variables. The lack of random assignment is the major weakness of the quasi-experimental design with respect to establishing causality (Harris et al., 2006). Even if there is a statistically significant difference in the outcome after an intervention, this statistical difference does not necessarily translate into causality (Harris et al., 2006). Therefore, it is
always prudent to ask a question about the possibility of alternative explanations for the apparent causal association (Harris et al., 2006). To make the evidence of causation more convincing it is important to address potential threats to validity and reliability. In this study, it has been done with respect to the threats that were regarded as relevant to the quasi-experimental study design and the interrupted time series by Shadish et al. (2002). In addition, Wagner et al. (2002) asserted that “time series designs are the strongest, quasi-experimental designs to estimate intervention effects in non-randomized settings “(p. 308), such as the setting for this study. Harris et al. (2006) noted that an advantage of the interrupted time series is that “with multiple measurements both pre- and postintervention, it is easier to address and control for confounding and regression to the mean” (p. 21). Statistically, according to Harris et al. (2006), the interrupted time series analysis has “a robust analytic capability, and there is the ability to detect changes in the slope or intercept as a result of the intervention in addition to a change in the mean values” (p. 21). Therefore, it can be said that, although a considerable effort was made to select the strongest design possible under the circumstances and to control for various threats to validity and reliability, caution should be exercised when establishing causality based on the findings of the study.

Furthermore, causal inferences need to be generalized with respect to (1) the constructs associated with specific persons, settings, treatments, and outcomes as measured and (2) the extent “to which the causal relationship holds over variations in persons, settings, treatments, and measurement variables” (Shadish et al., 2002, p. 341). According to Shadish et al. (2002), the formal random sampling method is the best way to ensure robust generalized causal inference. The sampling method in this study does not meet this standard. Accordingly, the need to exercise caution when making generalized causal inference from the study results is yet another
limitation of this study.

In summary, the limitations discussed in this section may have made the rationale for cause-and-effect-relationships among the variables and for further generalized causal inference less robust compared to a study based on a larger randomized sample with a larger number of observations allowing for a higher degree of control over any potential confounding factors. However, a thorough description of limitations should be an integral element of any research study to allow for appropriate theoretical and practical implications.

**Conclusions**

The purpose of the present study was to evaluate the effects of the change in the screening methodology used for placing kindergarten ELLs either in L1 or L2 reading instruction. This section will address conclusions based on the study results as outlined in Chapter 4 in relation to the two research questions that guided this investigation. The implications of these conclusions will be addressed in the later sections.

The first conclusion pertains to the first research question addressing the effects of the change in the screening methodology for placing kindergarten ELLs either in L1 or L2 on the reading achievement of ELLs as demonstrated by the test scores on the standardized OAKS reading achievement test. Based on the analyses of the study findings it can be concluded that the change in the screening procedure had a moderate positive effect on the reading achievement of all ELLs regardless of the instructional language that was used for their reading instruction. This effect was demonstrated in the statistically significant change in trend from the pre- to post-intervention period. However, no statistically significant difference was detected with respect to the means of test scores.

The second conclusion was made with regard to the second research question positing
that the effect of the change in the screening methodology would be more significant with respect the ELLs who received L1 reading instruction compared the effect on the group that received L2 reading instruction. Once again, the analyses of the study findings made it possible to conclude that the effect of the change in the screening procedure was more significant on the reading achievement of the ELLs who received L1 reading instruction than on that of the ELLs who received L2 reading instruction. However, the statistically significant difference was observed only in terms of the post-intervention slopes for the L1 and L2 groups, and not in terms of the means of test scores.

In summation, the change in the screening methodology had a positive effect on the reading achievement of ELLs as measured by the test scores on the third grade standardized reading achievement test with this effect being more significant with regard to the reading achievement of ELLs who received L1 reading instruction. The next section will discuss the conclusions proceeding from the study findings relative to the existing body of research.

**Discussion**

Despite the limitations of the study, its findings demonstrated that paying attention to the initial level of L1 proficiency when making decisions about placing kindergarten ELLs either in L1 or L2 reading instruction can have a statistically significant positive effect on their later English reading achievement as measured by the test scores on the OAKS standardized reading achievement test. This positive effect was specifically discernible with respect to those ELLs who received L1 reading instruction.

A review of the existing literature demonstrated that in recent years there has been general agreement about the fact that bilingual reading programs can have a positive effect on English reading achievement of ELLs (e.g., Mathes et al., 2010; Rolstad et al., 2005; Slavin &
Cheung, 2005). Therefore, the conclusions drawn from the findings of this study about the potential benefits of bilingual instruction for Spanish-speaking ELLs provide support for similar conclusions made in earlier researcher studies.

By contributing to the body of research acknowledging the benefits of bilingual reading instruction, this study also contributed to other areas of research related to bilingual education. First, because the bilingual program that provided the context for the study is a transitional bilingual program, the results of the study can help develop a better understanding of the role TBE can play in promoting English literacy in ELLs as a specific category of bilingual programs. According to Slavin et al. (2011), literacy instruction in L1 is the “hallmark” of such programs. The foregoing observation makes the context of the study and its results particularly relevant with respect to TBE. The literature review did not identify any research studies specifically focused on TBE as a distinct category of bilingual programs with the exception of a multi-year randomized evaluation of such programs by Slavin et al. (2011). The findings of the study by Slavin et al. (2011) failed to support an argument made by the proponents of TBE to the effect that L1 literacy instruction can support English reading achievement of Spanish-speaking children. Although the design and the goals of the present study were markedly different from those of the study by Slavin et al. (2011), one should not ignore the fact that, contrary to Slavin et al. (2011), the findings of this study endorsed potential benefits of TBE for Spanish-speaking ELLs.

Slavin et al. (2011) as well as some other authors (e.g., August & Shanahan, 2006; Calderón et al., 2011) argued that schools may choose to teach ELLs to read either in L1 or L2 for a variety of reasons, including cultural, political, and economic considerations. According to Slavin at al. (2011), what really matters is the quality of instruction and not the language of instruction. Based on this statement, quality of instruction and language of instruction may be
viewed as unrelated or even mutually-exclusive options for improving reading achievement of ELLs. In addition, among various considerations for choosing to teach ELLs to read in L1 Slavin et al. (2011) neglected to mention an academic reason. One can hardly argue with the fact that quality of instruction is essential for student academic success. Among other methods, quality of instruction may be enhanced by approaching the language-of-instruction issue from an academic perspective and using it as a tool to boost the reading achievement of those ELLs who may benefit the most from this instructional approach. Ultimately, all instruction, regardless of the language used for its delivery, should be of high quality. It is also safe to argue that placing kindergarten ELLs in a group where reading instruction is provided in the language in which they are currently more proficient can only help enhance instruction delivery. In addition, receiving reading instruction in their stronger L1 can help young ELLs to avoid an initial “failure experience” that may occur when ELLs instructed in their weaker L2 upon school entry.

Rosenberg and Fisher (2007) stated that differentiating instruction is an effective tool for facilitating the literacy and academic development of ELLs. Teachers should differentiate instruction by acknowledging student differences first and then tailoring their instructional practices accordingly (Rosenberg & Fisher, 2007). The fact that ELLs share a common L1 does not mean they are all equally proficient in that language. Therefore, one way to differentiate instruction is to take into account the level of L1 proficiency. The results of this study indicated that this approach can make a difference with respect to student reading achievement, particularly for those ELLs who had an opportunity to capitalize on their currently stronger L1. Quality of instruction does matter in terms of student learning outcomes, and it can be increased through differentiating instruction. From this perspective, the conclusions of this study do not come in direct contradiction, but rather compliment the conclusions of those researchers who
argue that the issue of the quality reading instruction for ELLs takes precedence over the issue of the language of instruction. The conclusions of this study clarify and expand on the importance of the quality of instruction by viewing the language of instruction as an additional tool for differentiating and thereby enhancing reading instruction of elementary Spanish-speaking ELLs.

By confirming the fact that the change in the screening methodology for placing ELLs either in L1 or L2 reading instruction can have a positive effect on the reading achievement of ELLs, the study supported the position of those researchers (e.g., Bernhardt, 2003; Escamilla, 2009; Hammer et al., 2011) who ascertain that literacy development in ELLs should not be viewed from a monolingual perspective, as well as the position of the authors who supported the need for a specific approach to the literacy instruction of ELLs (e.g., Bernhardt, 2003; Manyak, 2007). The results of the study illuminate the fact that L2 literacy learning for ELLs may be influenced by the level of their L1 proficiency and consequent ability to transfer literacy skills from L1 to L2. This premise highlights the importance of accounting for an existing L1 when thinking about how to support L2 literacy development. In this, the results of the study contribute to refuting claims about the redundancy of specialized approaches to the reading instruction for ELLs (e.g., Fitzgerald & Noblit, 1999, 2000; Hedgcock & Atkinson, 1993).

Although the study was not designed to specifically examine if L2 reading achievement of ELLs who were initially placed to receive L1 reading instruction was significantly different from that of ELLs who had English-only reading instruction, the results of study are nevertheless relevant with respect to an ongoing discussion about the role of the language of reading instruction for ELLs in general and on the efficacy of L1 reading instruction in particular as reflected in an extensive body of earlier research (e.g., Calderon et al., 2011; Domínguez de Ramírez & Shapiro, 2007; Mathes et al., 2007; Slavin & Cheung, 2005). Albeit not directly, the
study results support the position that the language of instruction is an essential component of the general discussion on how to boost reading achievement of ELLs (e.g., Domínguez de Ramírez & Shapiro, 2007) and that reading programs involving students’ L1 can be more effective than programs solely based on L2 (Rolstad et al., 2005; Slavin & Cheung, 2005;).

As was noted earlier, another important area of research in the field of bilingual education is represented by the studies on CLT (e.g., Atwill et al., 2007; Gabrielle et al., 2009; Proctor et al., 2006). This study did not attempt to explore the extent of CLT either in terms of isolated reading skills or in terms of the overall reading competency. However, by underscoring the role of initial L1 proficiency with respect to the later English reading achievement of ELLs receiving L1 reading instruction, this study contributed to the body of knowledge on the concept of CLT by confirming that L1 proficiency may influence the extent of CLT as it was demonstrated by the study results. In other words, a statistically significant difference in the reading achievement of ELLs as reflected in the trends before and after their L1 proficiency was taken into account confirmed that (1) CLT of reading skills from L1 and L2 is conceivable, and (2) the extent of CLT may increase with increased L1 proficiency.

To summarize, the results of the study relate to a number of issues addressed in research studies aimed at a more in-depth understanding of L2 literacy of ELLs. These issues include L2 literacy development in ELLs as opposed to native English speakers (i.e., bilingual approach vs. monolingual approach), and the need for specific approaches for L2 reading as opposed to using strategies proven effective with native speakers. In addition, the results of this research supported the premise that the language of instruction has its place in the discussion on how to assist ELLs to become competent L2 readers, and that L1 can be a language of choice for effective reading instruction of ELLs. Finally, the results of the study further endorsed the
concept of CLT and the premise about possible association between the level of L1 proficiency and the extent of CLT from L1 to L2. Guided by the foregoing summation, the next two sections will address theoretical and practical implications of the present study.

**Theoretical Implications**

This section will explore how the results of the study might add to or modify an understanding of the theoretical constructs that became the cornerstones for the theoretical rationale supporting this study. These constructs were primarily developed by Cummins and are explained in detail in Chapter 1. In addition, theoretical implications will be discussed from the research design perspective.

It has been about three decades since Cummins developed his theoretical constructs related to bilingualism and bilingual education. As was observed from the existing body of literature, most researchers who conducted studies in these areas used the constructs by Cummins as the theoretical rationale for their studies. The constructs most commonly cited as the foundation for the theoretical rationale are the principle of linguistic interdependence and the concept of CLT resulting from it (e.g., Cárdenas-Hagan et al., 2007; Manis et al., 2004; Proctor et al., 2006; Sparks et al., 2008). To varying degrees, the findings of many studies supported the theoretical propositions by Cummins (1979) that L1 and L2 in bilingual subjects are interdependent and that knowledge and skills acquired in L1 can be transferred to L2 (e.g., Atwill et al., 2007; Gabriele et al., 2009; Sparks et al., 2008). The fact that the results of the present study demonstrated a statistically significant difference in the reading achievement of ELLs who received L1 and L2 reading instruction as measured by the scores on the standardized reading achievement test can be also viewed as supporting the principle of linguistic interdependence and the concept of CLT. In other words, it is safe to say that ELLs who were
first taught to read in Spanish, their dominant language at the time, were able to not only develop reading skills in Spanish but were also able at a later time, when their L2 became stronger, to apply their L1 reading skills to support their reading in L2. In this, the study results added to and strengthened the body of research validating the theoretical construct by Cummins. This can be considered as a relevant theoretical implication in view of “the importance of generating and evaluating predictions from a coherent theory” (Cummins, 1988, p. 17).

With respect to TBE, Cummins (1988) noted that it was essential to acknowledge “the consistent patterns that do emerge in the research data regarding transfer or interdependence of cognitive academic skills across languages” (p. 17). Given the paucity of research on TBE and the fact that the results of a recent study on TBE by Slavin et al. (2011) did not demonstrate that initial L1 reading instruction supported L2 reading of ELLs, the next theoretical implication from this study is that its results can contribute to supporting the existence of linguistic interdependence with respect to TBE.

Although many researchers used the principle of linguistic interdependence and the construct of CLT to develop the theoretical rationale for their research as well as to validate them, one construct by Cummins (1979) that was largely ignored in research studies on bilingual education is the threshold hypothesis. The threshold hypothesis posits that the level of linguistic competence can affect the extent of transfer from L1 to L2. Although Durgunoğlu et al. (1993) pointed out that more research was needed to find out how L1 proficiency may influence the transfer of literacy skills from L1 to L2, to date, there appears to be no research specifically devoted to this issue. The statement by Durgunoğlu et al. (1993) about the role of initial L1 proficiency with respect to the extent of CLT is particularly relevant in the discussion of the methodology used for placing ELLs either in L1 or L2 reading instruction. As argued by Atwill
et al. (2007), the “one-size-fits-all” approach cannot guarantee reading success of all ELLs. However, there appears to be no extant studies where researchers set a specific goal to test the threshold hypothesis via examining how ignoring or taking into account the level of L1 linguistic proficiency may affect CLT. The present study was specifically designed to test the threshold hypothesis by examining the effects of the change in the screening methodology for placing kindergarten ELLs either L1 or L2 reading instruction. The essential element of the change was in modifying the screening procedure by including L1 proficiency criterion. In other words, the essence of the change was in taking into account the threshold hypothesis and assuming that, in terms of L2 reading achievement, all ELLs would benefit from this more intentional approach to placing them either in L1 or L2 reading instruction depending on their level of L1 proficiency. Based on the threshold hypothesis, it was also posited that ELLs with a more developed L1 would benefit the most. The results of the study confirmed the foregoing assumptions, particularly with respect to students in the L1 group. This special attention to testing the threshold hypothesis constitutes the third theoretical implication of this study.

In addition to examining the theoretical implications of the study with respect to the theoretical constructs it would be appropriate to examine theoretical implications from the perspective of the study design, that is, the use of the short interrupted time series analysis. While this analysis is widely used to study the effects of interventions in a variety of settings, such as healthcare (e.g., Wagner et al., 2002) and behavioral science (e.g., Ramsay et al., 2003), it is clearly underused in education research. The studies by Bloom (2003) and by Wong et al. (2009) are two of the few studies that used the interrupted time series approach to evaluate the effects of an intervention in the education setting. The fact that this study used a short interrupted time series design to examine the effect of an intervention in the school setting
advances the use of this design in education research. This is the fourth theoretical implication of the present research study.

In summary, the study addressed the need for more research or added to the existing research in several areas. First, similar to multiple other studies, this study contributed to the validation of the theoretical constructs by Cummins in the field of bilingualism and bilingual education, in particular, the principle of linguistic interdependence. Second, the study helped in establishing a pattern in the research findings on the existence of linguistic interdependence specifically with respect to TBE as a defined category of bilingual programs. Third, the results of the study supported the threshold hypothesis, thereby illuminating the importance of this theoretical construct in research on bilingual education. Finally, the study advanced the use the interrupted time series design to examine the effect of an intervention in the education setting.

**Practical Implications**

The idea for the study emerged from a practical need to understand the effects of the change in the screening methodology on the reading achievement of ELLs at an elementary school. This idea was one objective in a continuous quest for effective instructional approaches to boost English reading performance of ELLs. The effect of the change in the screening methodology was considered particularly pertinent with respect to the ELLs in L1 reading instruction because the intent of the NLL program is to support L2 reading development in ELLs. This student category has consistently demonstrated lower test scores on the state mandated standardized OAKS reading test. First, the No Child Left Behind mandate and now the Common Core Standards adopted by most states set increasingly challenging goals for student reading achievement. Therefore, the school finds itself under continuous pressure to demonstrate better test scores on the standardized reading achievement test.
Based on the size of the ELL population in American public schools, it is a safe assumption that the foregoing scenario (with the exception of a change in the screening procedure) is applicable to many elementary schools across the United States. Although what ultimately matters is student reading ability that is critical for mastering content in various subject areas, the challenge of reading underperformance of Spanish-speaking ELLs as demonstrated by the test scores continues to be a reality that schools across the United States have to deal with. Although this study has a number of limitations and its results should be generalized with extreme caution, it can offer several practical recommendations to teachers and administrators at elementary schools with a substantial population of Spanish-speaking ELLs and have or are considering a bilingual program. These recommendations for practice may be advantageous in helping Spanish-speaking ELLs to become more competent English readers.

The first recommendation for practice is related to the importance of taking into account the level of L1 proficiency when placing ELLs either in L1 or L2 reading instruction. To date, there is an absence of research on the methodology for developing and applying placement criteria that can be used for enrolling ELLs in bilingual programs. Durgunoğlu et al. (1993) observed that placement in bilingual programs is often based on the level of L2 proficiency and staff judgment. It appears that this situation remains in effect. However, the threshold hypothesis by Cummins (1979) posits that the level of L1 proficiency can act as a mediating factor with respect to CLT from L1 and L2. The results of this study supported the threshold hypothesis and demonstrated that the level of L1 proficiency should not be ignored in the process of placing young Spanish-speaking ELLs either in L1 or L2 reading instruction.

The second practice recommendation emerging from the results of the present study is related to the fact that educators should not assume that all ELLs will equally benefit from
participating in a bilingual reading program. The fact that a child comes from a home where a language other than English is spoken should not automatically imply that this other language is the child’s dominant language. Cummins (1984) maintained that children may appear to be fluent in a language at the basic communicative level, but are not able to rely on this language when completing more cognitively demanding tasks. Consequently, it is essential to take into account the distinction between BICS and CALP as different types of language proficiency (Cummins, 1981) when assessing the levels of language proficiency both in L1 and L2. From this perspective, it is advisable to pay careful attention to the selection of the screening measure used for this purpose. It is essential to not only choose a measure with proven validity and reliability but also to make sure that this measure assesses the level of CALP. Woodcock and Muñoz-Sandoval (1993a) argued that “the choice of an instrument used to determine eligibility for bilingual services has important programming ramifications” (p. 3). This position is based on the view by Cummins (1984) to the effect that language proficiency measures solely focused on the assessment of BICS can be misleading with respect to a subject’s ability to perform more cognitively demanding tasks with the use of this language. In addition, as pointed out by Durgunoğlu et al. (1993), young children may not have a strong grasp of their native language due to the lack of home support of their early language development. Therefore, placement decisions with respect to young ELLs should be made on a case-to-case basis taking into account the levels of language proficiency in terms of CALP both in L1 and L2.

Further, although the results of the study by Slavin et al. (2011) did not support the premise about the benefits of such programs, TBE should be viewed as a viable option for fostering English literacy achievement of ELLs. By focusing on reading, TBE addresses a fundamental competence that students need to master in lower grades so that they can use it
higher grades for content learning. In other words, to begin with, students need to learn to read, so that later they can read to learn. The transition from learning to read to reading to learn usually takes place between third and fourth grade. The timing of this transition coincided with the completion of early-exit TBE programs. The fact that the results of this study revealed a statistically significant positive difference in the reading achievement of ELLs placed in L1 reading instruction and that of ELLs placed in L2 reading instruction based on L1 and L2 proficiency screening (as measured by test scores on the third grade OAKS reading achievement test) confirms that early-exit TBE programs with a theory-based approach to student placement can foster the ability of ELLs to read in English so that they are better prepared to transition from learning to read to reading to learn.

Finally, as the results of current study support the premise about the benefits of bilingual education for ELLs, the implications for practice should include various measures aimed at promoting and supporting bilingual education. These measures could include eliminating the restrictions on the amount of time ELLs can be taught in their primary language (Slavin et al., 2011). In addition, more needs to be done to promote potential benefits of bilingualism and bilingual education for the general public and various education stakeholders, such as policy makers, administrators, teachers, parents, and students from the national to the state and local levels. More attention should be paid to the development and implementation of teacher training programs focused on bilingual education coupled with enhanced credentialing requirements for teachers working in bilinguals programs. Professional development opportunities for this category of teachers should also be expanded. It could also be advantageous if additional funded research on bilingual education generated more recent research data to be used in teacher preparation and professional development.
In conclusion, lagging reading skills of ELLs is a serious concern for many elementary schools with a significant population of Spanish-speaking ELLs because of the reading underachievement of this student category. In addition, such schools have been experiencing growing pressure in the form of increasing accountability and higher benchmarks. TBE programs can be a way to address this concern. However, careful attention should be paid to the criteria used for placing ELLs either in L1 or L2 reading instruction. These criteria should account for language proficiency not only in L2, but, most importantly, in L1 and utilize language proficiency screening measures designed to assess the level of CALP. The foregoing practical implications of this study are congruent with the constructs by Cummins that provided the theoretical rationale for L1 reading instruction and the role of initial L1 proficiency as well as the conclusions made by multiple research studies about the benefits of L1 literacy instruction.

**Recommendations for Future Research**

Findings from the current investigation raise a number of questions related to future research. Although a considerably body of research on bilingual education and bilingual reading instruction has been accumulated over the recent decades, many essential questions remain unanswered and full consensus on the merits of L1 reading instruction has not been achieved. In their 2005 meta-analysis of TBE programs, Slavin and Cheung concluded that more qualitative and quantitative research was needed to illuminate specific conditions that make L1 reading instruction more effective as well as to explain these conditions. This study addressed a specific condition, namely screening methodologies to be used for placing kindergarten ELLs either in L1 or L2 reading instruction. The results of the study suggest several possible directions for future research that could help to gain a better understanding of conditions pertinent to the efficacy of bilingual reading programs and fill in existing gaps in literature.
The first recommendation is to explore in greater depth placement methodologies (or lack of thereof) for placing ELLs in bilingual programs as opposed to placing them in English immersion programs. Zirkel (1974) cautioned that “without a simple but systematic method for determining and depicting language dominance, bilingual programs run the risk of becoming “dumping grounds” for pupils characterized by behavioral and/or learning difficulties rather than linguistic differences” (p. 8). Zirkel (1974) rationalized that “pupil placement by summary teacher opinion is subject to the inaccuracies of unconscious attitudes or skewed knowledge” (p. 8). Although the statements by Zirkel (1974) were made four decades ago, it appears that there has been no discussion in the academic literature on bilingual education about how student placement in bilingual programs should be approached. This study made an attempt to draw attention to this issue. However, results from both quantitative and qualitative research would be needed to start a discussion about placement from the theoretical and practical perspectives.

The second recommendation is related to the use of the interrupted time series analysis in education research. As pointed out earlier, this research approach has been extensively used in other settings, but it has been mostly neglected by researchers in education. At the same time, many authors underscored the value of interrupted time series analysis for quasi-experiments (e.g., Taljaard et al., 2014; Wagner et al., 2002; Wang, Walter, Grzbieta, & Olivier, 2013). Therefore, this type of analysis can offer a number of definitive benefits for education researchers. First, according to Taljaard et al. (2014), a major strength of this approach is in the ability to detect and evaluate intervention effects. In education settings, there is often a need for such evaluations. Next, Wagner et al. (2002) argued that the interrupted time series is the strongest quasi-experimental design. This statement would be particularly pertinent to education research where experimental studies may not be feasible for a variety of reasons. This study
demonstrated the use of interrupted time series with respect to an education intervention examined within the framework of a quasi-experiment. There clearly may be many other options for an effective use of this approach in education research.

Although in their overview of bilingual education García and Homonoff Woodley (2009) claimed that transitional bilingual programs are the most popular type of bilingual education in the United States, there appears to be a marked paucity of recent research examining such programs from the perspective of how TBE can foster English reading achievement of ELLs. The current study focused on a specific aspect of TBE, that is, the effects of the change in the screening methodology on the reading achievement of ELLs as measured by the test scores on the standardized reading achievement test. However, it would beneficial for all stakeholders in education if more recent and broader research data on how TBE can help to bridge an existing gap in the reading achievement of Spanish-speaking ELLs were available.

In addition, future research should consider longitudinal studies following students over time through their entire school career. In-depth case studies of TBE programs to gain better insights into the curriculum and teacher and student perceptions as well as true comparison studies of several schools with and without bilingual programs would also be advisable.

In summary, in light of the existing body of research both from the perspective of research topics addressed in education research and research design options used in research studies in general, several possible directions for future research can be pursued: (1) research on the placement methodologies for bilingual programs, (2) research in education settings with use of the interrupted time series, and (3) research related to reading instruction of ELLs in TBE.

**Final Comments**

Reading and literacy have caused some of the most controversial debates in education,
and there are numerous legal and political issues raised in this debate (Lipson & Wixson, 2007). However, the most important issue is to help all students learn to read and successfully use this competency for academic, professional, and recreational purposes. Educators, researchers, policy makers, and parents have long been looking into various ways to achieve this goal. Over several recent decades, a dramatic increase in the numbers of ELLs in public school in the United States and English reading difficulties experienced by these students only helped to make the debate more heated. The statistics with respect to the size and growth rates of the ELL student population, on the one side, and their reading underachievement and resulting academic difficulties, on the other side, are eloquent about the scope of this problem. When looking into possible solutions for the foregoing problem it is important to bear in mind that no one single solution will work for every child whose native language is other than English.

Therefore, it is necessary to scrutinize and try out various strategies when looking for an optimal match between individual students and the type of reading instruction. Initial reading instruction in students’ first language is a viable, albeit still debatable, option. However, L1 reading instruction cannot be expected to work equally well for all ELLs. Epstein (1977) argued that one of the crucial problems experienced by bilingual education is the lack of research showing that this education works. An extensive body of research literature has been accumulated since Epstein (1977) made this assertion. However, to date, there is still no full consensus about or in-depth understanding of how bilingual education can work best.

One such area is placement in bilingual programs. Four decades ago, Zirkel (1974) commented on the fact that “the need for a more efficient and effective means of assessing the degree of bilingualism for initially screening and placing” (p. 8) students in bilingual programs remains unfulfilled. Over the years, considerable progress has been made in terms of theory,
(e.g., seminal works by Cummins) and practice (e.g., development of language proficiency screening and assessment measures) related to this issue. However, the matter of how theory and assessment tools should be applied to the practice of student placement either in L1 or L2 reading instruction has not been given the attention due to it.

By defining its purpose as an evaluation of the effects of a change in screening methodology on the reading achievement of elementary ELLs, the current study made an attempt to fill this void. The results of the study demonstrated that focus on students’ L1 when placing them either in L1 or L2 reading instruction can make a difference in terms of their later reading achievement as measured by the scores on the OAKS standardized reading achievement test in third grade when students complete their L1 reading program and get ready to transition from learning to read to reading to learn.

Bilingualism is a complex and multi-faceted matter. Macnamara (1967) contended that “bilingualism is so complicated a phenomenon that one has the giddy feeling that in speaking of it one speaks of all things at once” (p. 5). Bilingualism can also be a considerable asset and provide ELLs with a material advantage if it can foster their cognitive growth by way of CLT from L1 to L2 and ultimately help them achieve much needed academic success on par with their English-speaking peers.
References


