January 2013

Predicting Academic Outcomes for Third Grade Students: Examining the Reading Achievement of Diverse Students Using the Diagnostic Lens of the Dynamic Indicators of Basic Early Literacy Skills

Zhivago Trevino Adderley
University of South Florida, zhivagoadderley@gmail.com

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the Educational Psychology Commons, Other Education Commons, and the Special Education and Teaching Commons

Scholar Commons Citation
Adderley, Zhivago Trevino, 'Predicting Academic Outcomes for Third Grade Students: Examining the Reading Achievement of Diverse Students Using the Diagnostic Lens of the Dynamic Indicators of Basic Early Literacy Skills' (2013). Graduate Theses and Dissertations.
http://scholarcommons.usf.edu/etd/4624

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Predicting Academic Outcomes for Third Grade Students: Examining the Reading Achievement of Diverse Students Using the Diagnostic Lens of the Dynamic Indicators of Basic Early Literacy Skills

by

Zhivago Trevino Adderley

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Psychological and Social Foundations College of Education University of South Florida

Major Professor: Kathy L. Bradley-Klug, Ph.D.
George Batsche, Ed.D.
John Ferron, Ph.D.
Lisa M. Lopez, Ph.D.

Date of Approval:
April 23rd, 2013

Keywords: Education Reform, Diverse Learners, English Language Learners, Specific Learning Disabilities

Copyright © 2013, Zhivago Trevino Adderley
Dedication

This manuscript is dedicated to several significant people in my life. First, this work is dedicated to the leading and influence of the Holy Spirit. What a guide! Thank you for directing me to Tampa and the University of South Florida to pursue graduate studies. Little did I know that Tampa would be the place to re-connect with my life’s partner, Romeika, where together we would fulfill our spiritual assignment to serve the body of Christ. Today (August 10, 2013) many witness the fulfillment of another one of your promises concerning me---to God be all the glory! Therefore, this work is dedicated to you Holy Spirit, the greatest teacher. May I remain your servant as you continue to guide my life.

Second, this work is also dedicated to my wife Romeika, my faithful and most ardent encourager, my dearest friend. Thank you for listening to the good, the bad and the ugly stories connected with this project. Also, thank you for being patient and believing in me even when I doubted. Now, it’s your turn; let’s go get your doctorate!

To my parents: Eardley and Irma Adderley-thank you! Thanks mommy for transferring your infectious faith, resilience and tenacious drive to me. I would not have completed this project if these characteristics were absent. I will forever honor you. I also honor my dad for quietly understanding that I was destined for more than a career in banking. Daddy, thanks for permitting me to dream….until we meet again. Fourth, this completed work is also dedicated to my parents through marriage: Franklin and Rovena Ferguson. Your wisdom in coaching me through to the finish line speaks well about your adult parenting skills—thank you too!
Finally, to my Rhaimee and my Jirehson---I give you the biggest air hugs as I say thank you! So many of the thoughts and reflections about diverse children and their families remind me that I must be the best daddy to you ever! I love you loads and loads and then some more!
Acknowledgments

I wish to express sincere thanks to the chairperson of my committee, Dr. Kathy Bradley-Klug for her advice, and support during this dissertation project. Dr. KBK: I am grateful for your decision to walk with me through the hard and challenging places that sometimes marked this journey. I am indebted to you for your unique demonstration of grace toward me. From the bottom of my heart “Thank you!”

Gratitude is also expressed to the remaining members of my committee: Dr. George Batsche, Dr. John Ferron, and Dr. Lisa Lopez. Your expert input, advice, guidance and re-direction also made the completion of this project possible. Finally, I acknowledge the long-term support of my friend, Dr. Kelly Powell-Smith who, at one time, was a member on this dissertation committee. Thanks for helping me get started and for quietly cheering me on to the finish line.
# Table of Contents

Table of Contents

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
</tbody>
</table>

## Chapter One: Introduction

<table>
<thead>
<tr>
<th>Overview of Educational Reform in the United States</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Reform in the State of Florida</td>
<td>5</td>
</tr>
<tr>
<td>Historical Context of Diverse Learners</td>
<td>7</td>
</tr>
<tr>
<td>Specific Learning Disabilities</td>
<td>9</td>
</tr>
<tr>
<td>English Language Learners</td>
<td>10</td>
</tr>
<tr>
<td>Linking Higher Achievement Standards to Authentic Measures</td>
<td>12</td>
</tr>
<tr>
<td>Rationale for Examining Diverse Students’ Achievement</td>
<td>15</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>16</td>
</tr>
</tbody>
</table>

## Chapter Two: Review of the Literature

<table>
<thead>
<tr>
<th>Overview</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Reform in the United States</td>
<td>18</td>
</tr>
<tr>
<td>The Improving Americas School Act of 1994</td>
<td>19</td>
</tr>
<tr>
<td>No Child Left Behind Act of 2001</td>
<td>21</td>
</tr>
<tr>
<td>A Blueprint for Reform</td>
<td>22</td>
</tr>
<tr>
<td>Accountability Reform in the State of Florida</td>
<td>24</td>
</tr>
<tr>
<td>Historical Context for Diverse Learners</td>
<td>26</td>
</tr>
<tr>
<td>Specific Learning Disabilities</td>
<td>28</td>
</tr>
<tr>
<td>English Language Learners</td>
<td>34</td>
</tr>
<tr>
<td>Context for Authentic Measures</td>
<td>44</td>
</tr>
<tr>
<td>The Dynamic Indicators of Basic Early Literacy Skills (DIBELS)</td>
<td>44</td>
</tr>
<tr>
<td>The DIBELS and State Assessments</td>
<td>47</td>
</tr>
<tr>
<td>DIBELS Measures and Diverse Students</td>
<td>58</td>
</tr>
<tr>
<td>Summary</td>
<td>63</td>
</tr>
</tbody>
</table>

## Chapter Three: Method

<table>
<thead>
<tr>
<th>Overview</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>65</td>
</tr>
<tr>
<td>Initial Student Data Retrieved</td>
<td>65</td>
</tr>
<tr>
<td>English Language Learners</td>
<td>66</td>
</tr>
<tr>
<td>Students with Specific Learning Disabilities</td>
<td>68</td>
</tr>
</tbody>
</table>
Appendix C: Confidentiality Agreement Form 154
Appendix D: ELL Reliability Check Form 155
Appendix E: SLD Reliability Check Form 156
List of Tables

Table 1  Study Sample: Third Grade Race/Ethnicity Membership by District  66

Table 2  Native Languages represented among ELLs in Districts A and C in 2006  68

Table 3  Demographic Data for the ELL and Non-ELL Students in the Data Set  68

Table 4  Study Sample: SLD Membership by Race/Ethnicity Category  69

Table 5  Definition of Specific Learning Disabilities per Federal Law, Florida Statutes, and the three Districts’ procedures  74

Table 6  Study Sample: Demographic Data of ELL and Non-ELL Groups by District  95

Table 7  Combined Descriptive Statistics of ELL and Non-ELL DORF observations  96

Table 8  Growth Curve Model: Prediction of DORF Estimates for English Language Learners  99

Table 9  Growth Curve Model: Variance Estimates for English Language Learners  99

Table 10  Race/Ethnicity Demographics of Students with Learning Disabilities by District  104

Table 11  Students with Learning Disabilities Descriptive DORF scores  104

Table 12  Growth Curve Model: Prediction of DORF Estimates for Students with Learning Disabilities  106

Table 13  Growth Curve Model: Variance Estimates for Students with Learning Disabilities  106
List of Figures

Figure 1  Predicted DORF trajectories for 16 Non-ELL and ELL groups 100
Figure 2  ELL scatter plot of residuals and predicted DORF scores 102
Figure 3  Graph depicting SLD DORF scores over time 107
Figure 4  SLD scatter plot of residuals and predicted DORF scores 108
Abstract

Despite the billions of dollars spent in the last forty years, America’s efforts toward closing the achievement gaps among diverse learners and their receptive counterparts have not been realized. Limitations noted in previous research discussed the need to examine the unique contributions of diverse learner variables as a way of determining their specific academic needs. The purpose of this study was to examine the intra- and inter-group growth trajectories of two diverse student groups (English Language Learners and Students with Specific Learning Disabilities) on reading achievement. The study employed a longitudinal, quasi-experimental research design utilizing archival data from 26,947 students’ files to answer two research questions. The first research question examined growth relationships between 3rd grade English Language Learner student categories on reading achievement while holding gender and socio-economic variables constant. The second research question explored the extent to which the initial levels and slopes of 3rd grade students with specific learning disabilities differed across racial and ethnic groups. Growth curve analyses were employed to answer both research questions.

Findings revealed significant intercept and slope relationships for the two groups on reading fluency measures. Significant differences were found between the reference group (i.e., Non-ELL females who were ineligible for free and or reduced lunch) and two of the ELL subgroups. The slope relationships were only significant for ELL students (ELL-LY) who were in the currently enrolled (i.e., receiving some type of ELL
instructional support or service) category. Gender and socio-economic variables were significant suggesting a negative influence on initial reading levels. Reading fluency (DORF) achievement findings relative to students with disabilities and their race and ethnic subgroups revealed White students’ initial DORF scores were significantly different from Hispanic and Black students’ scores. Race and ethnic slope variables were insignificant and homogeneous in nature. A discussion about these findings and their implications for closing the achievement gap for diverse students is provided in the document.
Chapter One

Introduction

Overview of Educational Reform in the United States

In 1983 the National Commission on Excellence in Education reported that the US was “a nation at risk” and would become disenfranchised if its educational system failed to equip students with “essential” literacy skills. The “nation at risk” report revealed that about 23 million adults were functionally illiterate despite legislative efforts that were implemented to reduce poverty and eliminate the achievement gap between White and disadvantaged students through the Elementary and Secondary Education Act (ESEA) of 1965. In 1992, data from the National Assessment of Education Progress’ (NAEP) still revealed huge disparities in reading achievement. Specifically, the NAEP findings suggested that only 25% and 37% of 4th and 8th graders, respectively, reached proficient levels in reading. These results were even lower for diverse student subgroups like Hispanics, Blacks and student with disabilities (Mullis, Campbell, & Farstrup, 1993). Reading proficiency was defined as the ability to read unfamiliar literal and inferential texts, make connections, and then draw conclusions about what was read.

The topic of education reform, especially as it relates to closing the achievement gap, remains a priority for federal and state legislators. A report by the U.S. Department of Education suggests that the US is regaining little ground relative to its top standing in academic achievement among first world countries, especially considering the increased population growth among diverse and lower income Americans. Recent data suggest that in the last twenty years, the population of White students in America’s schools declined by 13%, while the population of Hispanic students increased by 11%. Relative to income, data suggest that about 19% of families with school-aged children live in
poverty; an estimate that has increased from 15% in the last ten years (Aud et al. 2011; U.S. Department of Education, 2008). When presenting testimony to a hearing of the Committee on Health, Education, Labor, and Pensions, Andreas Schleicher (from the Organization of Economic Co-Operation and Development, France) posited that first world countries like Canada had small achievement gaps between advantaged and disadvantaged groups because these countries developed education standards that were focused and rigorous when compared to the inconsistent accountability standards that were developed by individual states within the US. He challenged further that the high stakes legislation known as No Child Left Behind Act of 2001 (NCLB) was ineffective because it only required states to assess students once a year. Schleicher referred to this as a “single bar” approach that unfairly identified too many successful schools as needing improvement (Schleicher, 2010).

Although high stakes testing is not a new concept to US policy makers or educators, it appears that its momentum increased during the last twenty years due, in part, to the influence of federal dollars that were attached to the two major reauthorizations of the ESEA (1965); in particular the Improving Americas School Act 1994 (Goals 2000) and NCLB reauthorizations that were signed into law. High stakes is a term used to describe consequences that are applied when schools do not meet minimum state requirements. These consequences include grade retention, diploma denial for graduating high school seniors, publication of school ratings, corrective action (e.g., local staff and or leadership re-assignment), and monetary incentives for those schools demonstrating progress (Secifert & Sutton, 2009; Zellner & Jinkins, 2001). Other issues surrounding the debate about high stakes testing and accountability are school
restructuring, ability grouping, inadequate academic progress of diverse and disadvantaged students, curriculum narrowing, variable graduation matrix, remedial or special education instruction, and state-to-state assessment reliability (Lips, 2008; Madaus & Clarke, 2001; Mead, 2007; U.S. Department of Education, 2008).

The Improving Americas Schools Act otherwise known as Goals 2000 required states to set reading standards with defined goals and benchmarks for all students. There also was a requirement for states to develop or revise education standards with clear and focused content to guide local curriculum development. Yearly assessments were to allow full participation for all students (including diverse learners) and to supply meaningful information about best practices by monitoring and evaluating the impact of state assessments on student learning and teacher instruction (U.S. Department of Education, 1998). Education reform activists argued that the loose standards written in the Goals 2000 legislation provided no specific path to increasing achievement because the law did not mandate how to hold states, schools, and students accountable. These realities might explain why many advocated for NCLB. The NCLB Act’s “…main purpose was to ensure that children had a fair, equal, and significant opportunity to obtain a high quality education….“ (p.1) (Florida Department of Education, 2004; NCLB, 2002). Based largely on reform efforts and national research committee results, the NCLB Act focused on four guiding principles: (1) stronger accountability, (2) increased local control, (3) more parent choice, and (4) proven teaching methods (US Department of Education, 2002; West & Peterson, 2003). Accountability measures were implemented using a framework called adequate yearly progress (AYP). AYP is the U.S. Department of Education’s approved measure for holding states accountable for
student achievement under NCLB. Specifically, states were required to set achievement baseline and benchmark targets each year (beginning in 2001/2002 school year) with special attention given to diverse student groups (e.g., Blacks or low income) where the goal was to progressively increase student achievement so that all students were 100% proficient in reading and math by 2014 (NCLB, 2002).

Current Department of Education secretary, Arnie Duncan, stated that the NCLB legislation impacted the US educational system in that it exposed the morally unacceptable achievement gaps between majority groups and diverse students; however, he argued for accountability standards that were fair, focused and flexible (Duncan, 2012). The proposed reauthorization of the Elementary and Secondary Act seeks to reward states and districts who make progress toward closing the achievement gap by (1) developing programs to increase teaching excellence, (2) encouraging curriculum and standards reform, (3) implementing and maintaining accessible longitudinal data that informs instruction, and by (4) ensuring that success for all students by turning around lowest performing schools where diverse students (e.g., Hispanics, Blacks and ELL) comprise higher percentages of the student population (Duncan, 2012; U.S. Department of Education, 2010a). Although the United States Congress has yet to pass the reauthorization of ESEA, the big ideas behind the Department of Education’s college and career ready standards and assessments were funded under the American Recovery and Reinvestment Act (ARRA) in 2009. In turn, the Department of Education developed the Race to the Top Fund, which is a competitive grant made available to states as a way of stimulating educational innovation and reform in America’s schools with the end goal of increasing student achievement and thereby ensuring better qualified college students and
career employees upon entry in these two respective arenas (U.S. Department of Education, 2009).

In sum, it appears the forty years of literature surrounding educational reform centers around the ESEA Act of 1965 and its major revisions. Theoretically, the tenants expressed in the legislations are that the educational system in the US needs to close its morally unacceptable achievement gaps between majority groups (e.g., Whites and non-disabled students) and minority or diverse groups (e.g., Hispanics and students with disabilities) if it is to regain its footing as a dominant global leader. Doing so will require clear, innovative, consistent, yet flexible education standards that are similarly aligned across all the states.

**Educational Reform in the State of Florida.** At the turn of the millennium, the Florida Legislature increased accountability for students and other school stakeholders by enacting a monetary rewards program and the A+ Plan for Education. The A+ Plan added two important pieces to already existing legislation. First, the plan required that all students demonstrate annual learning gains. Second, the plan categorized school performance and assigned a letter grade to every public school (Florida Department of Education, 2003). After ten years of reform efforts, data suggest that Florida is making gains relative to raising standards and closing the achievement gaps among students. Since 2005, Florida boasts of reading performances that have been slightly higher than the national average. Florida’s high school graduation rates (less than 70%) however, are below the national average of 75% (Aud et al., 2011; Aud et al., 2012; Perie, Grigg & Donahue, 2005; U. S. Department of Education, 2011a). Recently, Florida obtained a four-year Race to Top grant. The Race to the Top program is part of a larger federal
initiative called the America Recovery and Reinvest Act (2009) that provides competitive grants to states who, in turn, are encouraged to implement innovative educational reform that will result in higher student learning outcomes. Florida’s adoption of the Department of Education’s big idea on college and career ready standards and assessments include the goal of adopting rigorous formative assessments that will be used to identify and address student needs throughout the school year (U. S. Department of Education, 2012a).

Although credit is to be given to Florida for raising standards that are higher than 28 states in the US and for increasing its fourth grade reading proficiency rates from 21% in 1992 to 35% in both 2009 and 2011, these proficiency rates are still low. Significant gaps are also noted between Florida’s diverse learners and their comparison groups. For example, White students earn about 28.6% more points on reading achievement assessments than Black students; the achievement gap (i.e., 14.8%) between White and Hispanic students is not as large. The reading achievement divide is greater for English Language Learners (ELL) and non-ELL students (38.8%) and for students with and without disabilities (31.5%). In fact, as a group, diverse students in Florida fall in the lowest quartile of the NAEP results (Perie, Grigg, & Donahue, 2005; U.S. Department of Education, 2011b). In short, Florida, like much of the US, needs to refocus and re-examine its big ideas surrounding educational reform and determine how to use standards and high stakes testing, along with increasing teacher quality and accountability, to improve the current 35% fourth grade reading proficiency rate (Darling-Hammond, 1997; Dufour et al., 2004; Lips, 2008).
Historical Context of Diverse Learners

Clearly, the challenge of educating and increasing the achievement of diverse learners is huge, in part, because the definition of the diverse learner is evolving. Brantlinger (2003) asserted that diversity issues arose when there was a dominant group or culture and by default an “outside” or contrast group. Historically, educational data and assessment reports referred to the dominant group as suburban White students. Outsiders or contrasting groups were often considered as disadvantaged students because they lacked equal education opportunities; these students were often Black, American Indian or Hispanic Americans living in rural or urban areas (Kozol, 1991; Stewner-Manzanares, 1988; Walker De Felix, 1992). For example, in 1965 when the compensatory education or Title I programs Act was initiated (i.e., federal funding for high poverty schools and academically at-risk youth), disadvantaged students in rural and urban communities were the ones who typically qualified (Glass et al., 1970; U.S. Department of Health, 1969). Similarly, Title VII or the Bilingual Education Act of 1968 recognized and provided funding to school districts whose student population included Limited English Speaking Ability (LESA) students, now referred to as English Language Learners (ELLs) (National Council of Teachers of English, 2008; Stewner-Manzanares, 1988). Also, Coombs (1970) quantified disadvantaged American Indians as having lower test scores, lower percentages in high school completion and lower enrollment rates at the college level. Essentially, minority and or disadvantaged groups (e.g., Blacks, Mexicans, Puerto Ricans and American Indians) were identified by their low academic achievement, low socio-economic status, resident community and, in some places, by inequitable educational opportunities. Another demographic factor associated with disadvantaged
students prior to 1975 included US geographical regions such as the Southeast region where the achievement scores of these states often fell below the national average (Coleman, 1940; Coombs, 1970; Glass et al., 1970).

Post 1975, race, ethnicity, and socio-economic status (i.e., family income level and domicile demographics) were no longer the only categories that profiled the diverse learner in American public schools; diverse learners now included students with disabilities. Students with disabilities (i.e., mental and physical handicapping conditions) were considered disadvantaged or diverse because they were not given a “free and appropriate” education commensurate with their non-disabled peers prior to the passing of the Education for All Handicapped Children Act of 1975 (also known as Public law 94-142 or the Individuals with Disabilities Education Act as of 1990). Three additional purposes of P.L. 94-192 were to (1) provide support and funding to states and local agencies, (2) ensure that programs and services were unique thereby increasing the probability of student success, and (3) to periodically evaluate the success of programs and instruction given by local agencies (U.S. Department of Education, 2010b; Zettel, 1977). Currently, there are 6.5 million children and youth receiving services under the fourteen categories that are recognized and supported in the Individual with Disabilities Education Act (IDEA): Autism, Deaf-blindness, Deafness, Developmental Delay, Emotional Disturbance, Hearing Impairment, Multiple Disabilities, Orthopedic Impairment, Other Health Impaired, Intellectual Disabilities, Specific Learning Disabilities, Speech or Language Impairment, Traumatic Brain Injury and Visual Impairment (Aud et al., 2011).
Although the US has attained significant strides relative to students with disabilities as a diverse group, achievement data suggest that 75% of these students score below the mean on standardized assessments, 25% are perceived as not having the ability to keep pace with high school expectations, and overall, there is a significant learning gap among students with disabilities in core subject areas (i.e., reading, math, science and social studies) when compared to their non-disabled peers (U.S. Department of Education, 2010c; Wagner et al., 2003; Wagner, Newman, Cameto, Levine, 2006). This may explain why, of the fourteen categories of educational disabilities, some states report that students with specific learning disabilities (SLD) comprise about 50% of the special education services provided by local school agencies (National Joint Committee on Learning Disabilities, 2011). The next section of this chapter shifts the focus from a discussion about diverse learners in general to specifically providing an overview of two groups that comprise diverse students in America’s public schools: student with specific learning disabilities and English Language Learner students.

**Specific Learning Disabilities.** Thirty-eight percent (38%) of the 6.5 million students receiving special education services qualify under the specific learning disability category. A specific learning disability (SLD) is defined as a neuro-biologically based disorder of one of the cognitive processes that affect a student’s learning; it includes conditions of brain dysfunction or injury, but excludes learning difficulties that are primarily due to intellectual, emotional or physical handicaps (Individuals with Disabilities Education Act [IDEA], 1997; National Joint Committee on Learning Disabilities, 2011). A student qualified for SLD services when standardized assessment results revealed a significant discrepancy (i.e., unexpected underachievement) between
cognitive processes and academic achievement, and when the student demonstrated a need for special education. Although states have applied this definition when identifying students with learning disabilities, a common understanding and application across the US is absent (Individuals with Disabilities Education Act [IDEA], 1997; Kavale, Spaulding and Beam, 2009; National Joint Committee on Learning Disabilities, 2011).

**English Language Learners.** Hass and Huang (2010) states that English Language Learner (ELL) students are generally “…students in grades K-12 whose primary or first language is not English and who have not passed their states’ English language proficiency test” (p. ii). A more comprehensive definition was offered by Lopez (1995) who reported that ELL (formally referred to as Limited English Proficient) students typically came from homes that were culturally different, communicated in a language other than English, and demonstrated proficiency in their first language’s receptive and expressive abilities. Recent data suggest that there are 11.2 million (or 21% of the school-aged population) ELL children speaking over 400 languages in US schools. When disaggregated by race/ethnicity, Hispanic students (e.g., Mexicans, Puerto Ricans or Cubans) account for at least 70% and Asian students (e.g., Chinese, Japanese, or Filipino) account for 13% of the ELL student population (Aud et al., 2011). Clearly, ELL students, based on their ethnic composition alone, are considered a complex and heterogeneous group relative to culture, academics and respective behavioral profiles (Lopez, 1995; National Council of Teachers of English, 2008).

Relative to academic performance, the gap between ELL and non-ELL students remains wide, although improvements were noted since 1998. However, the NAEP results did not indicate any significant differences in reading between the 2005, 2007 and
2009 reports (Lee, Grigg, & Donahue, 2007; U.S. Department of Education, 2009). Reports indicate that ELL students score 37 points lower than their non-ELL peers on state mandated assessments. Conversely, only 2% of ELL students have achievement scores that fall in the highest quartile of state mandated assessments; in contrast, 24% of ELL students’ performance falls in the bottom quartile (National Center for Education Statistics, 2011). Federal initiatives to address underachievement among ELLs have included the Bilingual Education Act of 1968 and the English Language Acquisition Language Enhancement and Academic Achievement Act, 2002 also known as Title III; the latter was part of the NCLB initiative. The spirit of both legislative acts, as well as the re-authorizations in between, was to provide funding for programs (e.g., sheltered English immersion or dual language) that would facilitate positive academic results for English Speakers of Other Languages (ESOL) students (National Council of Teachers of English, 2008). Unfortunately, the current ELL ranking, mentioned above, along with the ESEA reauthorization proposal, suggest that these programs may have been ineffective in addressing the achievement gap (U. S. Department of Education, 2010b).

In summary, it is clear that the definition of the diverse learner has evolved over time. Pre-1975 definitions may have been sufficient with categories like Black, Hispanic, American Indian, poor, disadvantaged, rural or urban; however, the shift in the demographics and culture of American schools in the last 30 years, and the continued achievement gap between student groups (e.g., students with and without learning disabilities and ELL and Non-ELL students) suggest that a deeper understanding of the diverse learner is warranted if the challenging work of narrowing the achievement gap is to be realized.
The next section of this chapter provides an overview of how curriculum-based measures in reading are authentic assessments that are common, technically adequate, and can provide information about diverse learners and the general school population that are relevant to local school principals and or state or federal analysts in the Department of Education. Specifically, a brief overview of Curriculum-Based Measurement, Reading (CBM-R) and the Dynamic Indicators of Basic Early Literacy Skills, Oral Reading Fluency (DORF) are provided.

**Linking Higher Achievement Standards to Authentic Measures**

In light of the educational challenges raised by reform advocates about the morally unacceptable achievement gaps between various student groups, considerable attention has been given to the use and acceptance of authentic monitoring measures like Curriculum-Based Measurement (CBM) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). CBM involves the use of technically sound and relevant measures to identify skill gaps and inform instruction in core subject areas like reading and math. The use of CBM has increased as reform advocates argued that traditional measures did not provide enough information to develop, monitor or evaluate student performance (Deno, 1986). CBM was designed primarily for formative evaluation at the classroom level, but increasingly these measures were used to provide indicators to schools and districts about how subgroups within schools were performing (Fuchs, 2004; Marston, 1989; Shinn, 1995). The DIBELS were developed based on research and measures from CBM-R. Good and Kaminski, (2002a) developed the DIBELS measures because of research in the lower elementary grades suggesting that early identification and intervention with children at-risk for reading failure allowed for intervention to
reduce reading failure. Like CBM, DIBELS measures are linked to students’ curricula, are of short duration, technically adequate, inexpensive, and are sensitive to small improvements in students’ performance over short periods of time.

The moderate to strong criterion-related validity of CBM and DIBELS with state assessments are well documented and have been used to predict student performance on high stakes tests in several states (McGlinchey & Hixson, 2004; Millet, 2011; Stage & Jacobsen, 2001; Wiley & Deno, 2005). For example, McGlinchey and Hixson found a moderately strong relationship between oral reading rates and the Michigan Educational Assessment Program high stakes test scores. Similarly, researchers in Florida report moderate to strong correlations between DIBELS and FCAT reading measures (Buck & Torgesen, 2002; Young-Suk, Petscher, Schatschneider, & Foorman, 2010). Although strong correlations between high stakes state assessments and CBM or DIBELS measures are good, the advocates for authentic measures contend that identification of an at risk student is only one part of the problem solving process (Deno, 1986; Deno, 1989; Fuchs & Fuchs, 2006). Specifically, the DIBELS measures are designed to be used within a prevention-oriented Outcomes Driven Model (ODM) to assess early literacy skills. The model is prevention oriented and is focused on immediate responding to problems encountered during initial reading skills acquisition. Periodic assessment outcomes are obtained and linked to grade level benchmarks so that instructional decisions can be made regarding students’ response to instruction within the core curricula. Student performance is then determined as low risk, strategic or intensive; the latter classification representing the greatest need of support in order to reach the next benchmark (Fuchs, 1989; Kaminski, Cummings, Powell-Smith, & Good, 2008; Marston, 1989).
The outcomes driven model described above is an example of a multi-tiered support system (MTSS). MTSS, often referred to as response to intervention (RtI), uses data to inform research-based instruction, assessment, intervention and monitoring practices in order to evaluate students’ academic performance. Also, as indicated in its name, the MTSS framework is multi-tiered encouraging local education agencies to make decisions about students based on students’ response to instruction or intervention. Although there is some variation of the tiered support framework across school districts, greater levels of intensive support services are expected if students move from Tier I to Tier III. Tier I serves all students focusing on the school’s core curriculum, high quality instruction, and its universal screening measures (e.g., DIBELS) to ensure that at least 80% of the students meet expected benchmarks. Also, the MTSS process helps to identify and support students who are in need of Tier II support. Tier II support is implemented when students are (a) academically functioning at a level that is below their peers and or when (b) their rate of progress indicates that their learning gap is not closing in relation to the peer group. Tier II requires additional instructional/intervention support to change the direction of a student’s academic level and slope trajectory. Within the MTSS, Tier III discussions and decisions are driven by data that indicate limited response to high quality, research-based, intensive instruction and intervention resources that are implemented with integrity at Tier II (Ardoin, Witt, Connell, & Koeing, 2005; Batsche, Kavale, & Kovaleski, 2006; Fuchs & Fuchs, 2006; Kaminski, Cummings, Powell-Smith, & Good, 2008; National Center for Special Education Research, 2013).
Rationale for Examining Diverse Students’ Achievement

Despite the billions of dollars spent in the last forty years, America’s efforts toward closing the morally unacceptable achievement gap among diverse learners and their receptive counterparts has yet to be attained (Aud et al., 2011). Although unacceptable, Batsche et al., (2006) state that “…it is the gap that triggers referrals, and it will be the gap that will be used to determine intervention success” (p. 12). To this end, empirical research in the last decade demonstrates that the DORF measures provide moderate to strong predictive validity with high stakes state assessments that are considered complimentary progress monitoring measures that may be useful in closing the achievement gap (Baker et al., 2008; Buck & Torgesen, 2002; Shaw & Shaw, 2002; Wilson, 2005). However, one consistent limitation noted in the DIBELS literature is that there are few studies that have purposefully investigated the impact of authentic measures on diverse students’ subgroups (e.g., minorities, ELL and students with learning disabilities) over time. For example, a study by Baker et al. (2008) found that DORF slopes were a strong predictor of student performance, but the authors acknowledged that their study did not examine the unique contributions of reading slopes with diverse students or even with subsets of this growing student population. Also, the few studies that have examined the relationship between authentic measures and the performance of diverse students have suggested that intra-group investigations are warranted, as the reading slopes among these subgroups may be different, and as such may require different intervention models if success is to occur. Exploratory findings about the longitudinal impact of authentic measures on diverse subgroups further suggest that demographic variables (e.g., gender, SES, race/ethnicity) should be included in growth
curve models, as some of their unique contributions on reading trajectories are either consistent or tend to vary across and among learners (Artiles, Rueda, Salazar, & Higareda, 2005; D’Angiulli, Siegel, & Maggi, 2004; Kieffer, 2011).

**Statement of the Problem:**

In order to address the gaps in the literature, the present investigation examined “within” and/or between group trajectories for students with learning disabilities and English language learners. The results cast additional light on how authentic measures like the DORF predict diverse learners’ reading growth over time and, as a result, provide direction toward addressing students’ specific instructional needs. Using longitudinal archival data from three districts in Florida and the current version of the Statistical Analysis System software (SAS 9.3), the present study investigated the following questions:

1. To what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (ELLs) enrolled in an ESOL program (LY), ELLs not enrolled but monitored (LF), ELLs not enrolled and no longer monitored (LZ), and non-ELLs (ZZ) when holding gender and SES constant?

2. To what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnic groups?
Chapter Two

Review of the Literature

Overview

The following literature review discusses the state of education reform in the US and also summarizes some of the challenges and goals associated with increasing literacy while simultaneously closing an almost fifty-year reading achievement gap between advantaged and at risk students. The review begins with a discussion about the US as a nation “at risk” for educational failure and highlights the three most current legislations of the Elementary and Secondary Education Act (ESEA) of 1965. This chapter provides a summary of the major portions of the Improving America Schools Act (1994), the NCLB Act (2002), and the recently proposed Blueprint for Reform. Next, the review provides a brief history of the State of Florida’s accountability system. A historical overview about diverse learners will also be presented. Specifically, the overview will show how the definition of the diverse learner is evolving from identifiers of race, SES and residential geography to also include ELL students and students with disabilities. The chapter then transitions to a review about the research surrounding the “complimentary” nature of an authentic assessment measure (i.e., the DIBELS) with state assessments. Finally, the chapter examines the strengths and weaknesses of recent longitudinal research exploring the utility of the DIBELS with ELLs and students with learning disabilities.
Educational Reform in the United States

In 1983, the National Commission on Excellence in Education reported that the US was “a nation at risk” and would become disenfranchised if its educational system failed to equip students with “essential” literacy skills. Despite legislative efforts from the Elementary and Secondary Education Act (ESEA) of 1965, the “Nation at Risk” report revealed that about 23 million adults were functionally illiterate and that the achievement gap between White and disadvantaged students remained wide. The ESEA’s primary purpose was to provide Title I funds or financial support for children from low-income families in the form of programs (e.g., free and reduced lunch) and services. Legislators reasoned that the additional resources would even the playing field and, in turn, raise academic achievement. An inherent flaw in the ESEA legislation was that low achievement for America’s disadvantaged students was not an equity-based issue alone, but low achievement was also due to inconsistent educational standards and accountability among the states. This might explain why the National Commission on Excellence in Education recommended an educational system that included high expectations, goal setting and rigorous measurable standards. Also, the commission stated that if the level of academic mediocrity remained, then the majority of US citizens would be lacking in essential literacy skills and would find themselves disenfranchised in a global economy (Jennings, 2012; Kosters & Mast, 2003; U.S. Department of Education, 1983).

Although America responded to the Nation at Risk report, illiteracy among adult Americans, underachievement in America’s public schools, and general economic disenfranchisement remained in the 1990’s. For example, Wade (1998) reported that 30
million Americans (an increase of 7 million since 1983) were considered functionally illiterate when illiteracy was defined as the inability to make inferences or to demonstrate a literal understanding of what was read. Data from the National Assessment of Education Progress’ (NAEP) assessments revealed that the majority of 4th graders who did not demonstrate the ability to read and make inferences were often diverse or disadvantaged students such as Hispanics, Blacks, Indian American, or students from urban communities (Coombs, 1970; Mullis, Campbell, & Farstrup, 1993; National Center for Education Statistics, 2011).

**The Improving Americas School Act of 1994.** In addition to illiteracy and race equity concerns, America’s need to compete in a global economy also encouraged a shift toward accountability or outcome-based standards reform. The Improving Americas School Act of 1994, also known as Goals 2000 provided a “central” focus for the states where the big ideas surrounding learning were attached to quantifiable goals, supported by outcome measures. The big ideas of Goals 2000 were that key stakeholders in American schools would focus on (1) school readiness, (2) school completion, (3) student achievement, (4) professional development, (5) parental involvement and (6) school safety. Relative to quantifiable goals in student achievement, Goals 2000 mandated that all students demonstrate competency on challenging material and show increased academic performance. Similarly, states were encouraged to ensure that graduation rates increased to 90% by the year 2000. In an effort to encourage states to move toward a central focus, the National Education Standards and Improvement Council was charged with the responsibility of reviewing and certifying that state standards and student
outcomes were comparable or higher than standards and goals developed by the federal government (U.S. Department of Education, 1998).

Also, the Improving Americas School Act (1994) required that states change mandated assessment practices to (1) allow all students (including diverse learners) to participate; (2) inform key stakeholders about progress; (3) facilitate improved classroom instruction; (4) provide comparison achievement data, expectations and measures; and to (5) motivate students, schools and districts toward higher expectations. An important caveat in the Improving America’s School Act was the decision that assessment measures were not to be used by the states as prerequisites for graduation, grade retention or promotion for at least five years from the date of the enactment (U.S. Department of Education, 1998; U.S. Department of Education, 2000).

The fact that the caveat to delay consequences for five years was placed in the Goals 2000 legislation should have alerted educators that an era of high stakes accountability was on its way. High stakes is a term used to describe consequences that are applied when schools do not meet minimum state requirements. Some consequences have been grade retention, diploma denial for graduating high school seniors, publication of school ratings and monetary incentives for those schools demonstrating progress (Zellner & Jinkins, 2001). Other issues surrounding the debate about high stakes testing and accountability are school restructuring, inadequate yearly academic progress of diverse and disadvantaged students, curriculum narrowing, variable graduation matrix, remedial or special education instruction and state-to-state assessment reliability (Lips, 2008; Madaus & Clarke, 2001; Mead, 2007; U.S. Department of Education, 2008).
No Child Left Behind Act of 2001. The goal of No Child Left Behind (NCLB) was to change the outlook of America’s schools by ensuring that academic standards were in place in every school and that the key stakeholders were held accountable for those standards (West & Peterson, 2003). The NCLB Act focused on four principles: (1) stronger accountability, (2) increased local control, (3) more parent choice, and (4) proven teaching methods (U.S. Department of Education, 2002; West & Peterson, 2003). Stronger accountability for the states meant that state standards and assessment measures were submitted to the Department of Education for approval before federal funding was provided. Also, states were required to assess student performance in reading and math in grades three through eight. They were required to display the aggregated and disaggregated data to ensure that all students, especially those from disadvantaged groups (e.g., Blacks, Students with Disabilities) were making adequate yearly progress. Further, public dissemination of the data in the form of school “report cards” as well as reports on school safety were two additional ways that the NCLB ensured accountability. Schools and districts not making adequate yearly progress toward goals were considered “in need of improvement” and were subject to corrective action. For example, a corrective action could have included the implementation of a new and proven curriculum series along with the appropriate professional development. Districts and schools that made adequate or exemplary progress were rewarded.

Public school accountability was directly related to enhanced parental choice. That is, NCLB provided parents with additional school choice if their neighborhood school failed to meet state standards for at least two consecutive years. Additional school choice included the option of parents transferring their children to a better performing
public or charter school in the district, or the option of parents transferring their children to a private school using base student funding that was allotted to all students in a given district. In addition, NCLB stated that districts were required to provide supplemental educational services to students from low-income families if the schools did not meet state standards for three years.

The NCLB Act also allowed greater flexibility and decision-making at state and district levels on how federal dollars were used to address educational needs, especially in districts where diverse students were located. More specifically, districts had the flexibility of transferring up to 50% of any federal program fund to another federal fund if the district deemed it necessary. For example, a district might elect to take up to 50% of the Educational Technology funds to supplant funds that were needed to provide professional development to the district.

Relative to proven teaching methods, the NCLB Act established the *Early Reading First* and *Reading First* programs. Essentially, *Reading First* required schools receiving NCLB funding to use teaching practices (e.g., explicit instruction) and evidenced-based interventions that were research-based and effective. Moreover, *Reading First* focused its attention on sub-skills in reading proficiency (e.g., phonemic awareness, vocabulary, fluency) in grades K – 3 so that by the end of third grade, every child was reading on or above grade level (NCLB, 2002; Snow, 1998).

**A Blueprint for Reform.** Current US Secretary of Education, Arnie Duncan, states that the NCLB is to be credited for exposing the morally unacceptable achievement gaps between majority groups and diverse students. However, he argues that the flaws in NCLB warrant reform. The big ideas surrounding the proposed reauthorization of the
ESEA seeks to (1) make America college and career ready, (2) ensure equity and opportunity for all students, (3) ensure fair accountability, (4) meet the needs of diverse learners, (5) raise the achievement bar, and (6) promote continuous improvement in schools (Duncan, 2012; U.S. Department of Education, 2010b).

Relative to making America college and career ready, the current administration proposes to have states improve core standards (e.g., English, Math, and Science) and collaborate on standards that are common among the states. Also, states are encouraged to revise standards so that teachers and students have clear expectations about what they are required to know and do. Requirements are to be made available to parents to facilitate (a) where their child is in the curriculum, and to (b) allow parents to use the standards as an evaluation tool for schools. Guidelines supporting other content areas (e.g., History) are less rigorous. Equity efforts require states to ensure that high poverty schools receive comparable funding when compared to their low poverty counterparts. High poverty schools will also be given more flexibility on how to support disadvantaged students. Tied to the proposed standards reform is the big idea about rigorous and fair accountability support systems. Schools and districts will be rewarded for raising achievement scores and graduation rates, especially among low performing schools. Similar to NCLB, data are to be disaggregated so that information about at risk students (e.g., low SES, SWD and ELL students) is readily available to facilitate planning. Also, the Blueprint for Reform legislation addresses the need to develop and/or improve English language proficiency standards so that they are aligned with the revised core standards. The provisions are explicit so that states can better track ELL students’ progress, modify instruction and thereby increase ELL graduation rates.
The final two big ideas (i.e., raising the achievement bar and improving schools) will not be realized if “challenge” schools are not identified. Challenge schools are defined as having low graduation rates, high diverse or subgroup populations, and are underachieving academically. Volunteer states apply to the Department of Education and compete for Race to the Top funds (i.e., funding generated from the American Recovery and Reinvestment Act of 2009) in an effort to turn challenging schools around. The Race to the Top is a competitive grant made available to states as a way of stimulating educational innovation and reform in America’s schools with the end goal of increasing student achievement and ensuring that US citizens are college and career ready by 2020 (U.S. Department of Education, 2009; U.S. Department of Education 2010b).

**Accountability Reform in the State of Florida**

Florida’s reform has also been ongoing. In 1996, the State of Florida launched its Blueprint 2000 initiative for increased educational achievement; this corresponded with the federal government’s Goals 2000 initiative launched in 1994. Specifically, Florida’s Blueprint goals were to increase (1) school readiness, (2) graduation rates and readiness for the workforce, (3) student performance, (4) the efficacy of the learning environment, (5) school safety, (6) highly qualified teachers and staff, (7) adult literacy, and (8) parental involvement. By 1998, Florida implemented a series of revisions to facilitate higher student performance. Revisions included the establishment of school advisory councils and modifications to the Sunshine State Standards and its assessment correlate the Florida Comprehensive Assessment Test (Florida Department of Education, 2003).

The Florida Legislature also increased accountability measures for students and other school stakeholders by enacting a monetary rewards program called the A+ Plan for
Education in 1997. This program was first funded in 1998. The A+ Plan added two important pieces to the already existing legislation. First, the plan required that all students demonstrate one year’s gain in the curriculum in one year’s time. Second, the plan mandated increased accountability for student achievement in all schools. Specifically, students were assessed more often and needed to demonstrate learning gains in three of the basic skills areas (i.e., reading, mathematics and writing). Schools were assigned a letter grade based on learning gains across the student population and within specific disadvantaged subgroups (e.g., Black males or Hispanic students). Schools that demonstrated adequate learning gains were recognized and rewarded monetarily, while schools that received low or failing grades were penalized. School grades were first published in 1999 (Florida Department of Education, 2003).

Reform efforts have proven to be beneficial in Florida. NAEP data comparisons from 1996 to the present indicate that Florida’s reading achievement scores are improving. For example, the 1996 and 1998 NAEP fourth grade results revealed that Florida’s schools fell slightly below the national average in reading; however, results from the 2007 NAEP report indicated that Florida’s fourth grade reading scores were slightly above the national average. No significant improvements in fourth grade reading were noted in the last five years, however. It is also noteworthy that only 35% of fourth graders are at or above proficiency in reading (Lee, Grigg & Donahue, 2007; U.S. Department of Education, 2011a). When placed in the context of diverse learners, Florida’s statistics support the need for educational reform. For example, White students earn about 28.6% more points on reading achievement assessments than Black students and 14.8% more than Hispanic students. The achievement gap also is greater between
English Language Learners (ELL) and non-ELL students (38.8%), and for students with and without disabilities (31.5%). In short, the majority of diverse students fall in the lowest quartile on reading assessments (National Center for Education Statistics, 2011b; Perie, Grigg, & Donahue, 2005; U.S. Department of Education, 2011).

As educators and reform advocates have indicated the US, and specifically the state of Florida, needs to increase its attention on diverse student learners if 4th grade reading proficiency rates are to increase. Arguments about fair accountability, student progress, dynamic learning organizations, equity for all learners, and shared decision making are probably all factors that impact student achievement (Dufour, Dufour, Eaker, & Karhanek, 2004; Duncan, 2012; Jennings, 2012; Kosters & Mast, 2003; Lips, 2008; NCLB 2002; Walker de Felix, 1992). The present chapter now provides a brief historical context of diverse learners and then summarizes the research about two members of this group: students with learning disabilities and ELL students.

**Historical Context for Diverse Learners**

Historically, national or state assessment data referred to the dominant learning group as suburban White students. Outsiders or contrasting groups were categorized as disadvantaged students; these students were often Black, American Indian or Hispanic Americans living in rural or urban areas. For example, in 1965 when President L. B. Johnson initiated the compensatory education or Title I programs Act (i.e., federal funding for high poverty schools and academically at-risk youth), disadvantaged students in rural and urban communities were the ones who typically qualified (Glass, 1970; U.S. Department of Health, 1969). Essentially, minority and or disadvantaged groups (Blacks, Mexicans, Puerto Ricans and American Indians) were identified by their low academic
achievement, low socio-economic status, resident community and, in some places, by inequitable educational opportunities. Another demographic factor associated with disadvantaged students prior to 1975 included US geographical regions such as the Southeast region where states’ and districts’ achievement scores often fell below the national average (Coleman, 1940; Coombs, 1970; Glass, 1970).

Post 1975, race, ethnicity, and socio-economic status (i.e., family income level and domicile demographics) were no longer the only categories that profiled the diverse learner in American public schools. Students with disabilities (i.e., mental and physical handicapping conditions) were also considered disadvantaged or diverse learners because students with disabilities were not given a “free and appropriate” education commensurate to their non-disabled peers prior to the Education for All Handicapped Children Act of 1975 (also known as Public law 94-142). Three additional purposes of P.L. 94-192 were to (1) provide support and funding to states and local agencies, (2) ensure that programs and services were unique thereby increasing the probability of student success, and (3) to periodically evaluate the success of programs and instruction given by local agencies (U.S. Department of Education, 2010). Currently, there are 6.5 million children and youth receiving services under the fourteen categories that are recognized and supported under IDEA: Autism, Deaf-blindness, Deafness, Developmental Delay, Emotional Disturbance, Hearing Impairment, Multiple Disabilities, Orthopedic Impairment, Other Health Impaired, Intellectual Disabilities, Specific Learning Disabilities, Speech or Language Impairment, Traumatic Brain Injury and Visual Impairment (Aud et al., 2011).
Achievement data relative to students with disabilities suggest that 75% score below the mean on standardized assessments, 25% are perceived as not having the ability to keep pace with high school expectations, and overall, there is a significant learning gap among students with disabilities in core subject areas (i.e., reading, math, science and social studies) when compared to their non-disabled peers (Wagner et al., 2003; Wagner, Newman, Cameto, Levine, 2006; U.S. Department of Education, 2010b). This may explain why, of the fourteen categories of educational disabilities, students with specific learning disabilities (SLD) comprise 38% of the special education services provided by local school agencies across the US (Aud et al., 2011).

Specific Learning Disabilities. There are about 2.5 million students receiving special education services under the specific learning disability category (Aud et al., 2011; U.S. Department of Education, 2012b). A specific learning disability (SLD) is defined as a neuro-biologically based disorder of one of the cognitive processes that affect a student’s learning; it includes conditions of brain dysfunction or injury, but excludes learning difficulties that are primarily due to intellectual, emotional or physical handicaps (Individuals with Disabilities Education Act [IDEA], 1997; National Joint Committee on Learning Disabilities, 2011). Historically, a student qualified for SLD programs and or services when standardized assessment results revealed a significant discrepancy between cognitive processes and academic achievement, and when the student demonstrated a need for special education. On average, students with SLD score 20-points below the mean on standardized achievement tests, while only 11% score at or above the mean (IDEA, 1997; National Joint Committee on Learning Disabilities, 2011). Race/ethnicity data from this group suggest that White students outperform Hispanic and
African American students by 7 to 13 points (Wagner, Newman, Cameto & Levine, 2006). Early models of SLD focused on supplanting core instruction with small group or pull out one-to-one instruction and or related support services. Annual assessments were required to inform decision making and increase the probability of success (U.S. Department of Education, 2010b).

One example of early research that examined the effectiveness of programs utilized by students with disabilities was a study conducted by Mosby (1979). Mosby implemented a “curriculum on tape” intervention to provide support to students with reading problems in social studies content. The specific intent was to determine if audio-taping the reading portions of a social studies curriculum would increase academic achievement and decrease undesirable behaviors in the classroom. The intended participants were 50 7th graders from two junior high schools. Students were identified based on academic performance, but also students’ scores on intellectual and academic achievement test batteries (e.g., the Wechsler Intelligence Scale for Children and the Peabody Picture Vocabulary Test) were factored into the selection process. Direct teacher participants were two resource room teachers and their two assistants. Also, secondary teacher participants were general education social studies teachers.

The intervention was a multi-tiered accommodation process. Specifically, the intervention process required resource room teachers to liaise with content area teachers quarterly so that printed reading material and tests with multiple choice formats were transferred auditorally to a cassette tape. In turn, the social studies and other content area students accessed and listened to the cassette tapes in the school’s library. Students were also permitted to check out the tapes from the library in order to listen to the recordings at
home. Student participants were given the option of listening to the tape alone or listening to the tape while reading the text. Also, resource room teachers and their assistants were responsible for transcribing social studies students’ written responses if students demonstrated prior writing deficits. It was noted that social studies students were not allowed to be “pulled out” during the specific content instructional time. The final accommodation provided by the resource teachers was that they reduced the social studies assignments into small manageable chunks.

Mosby (1979) used the Stanford Achievement Test (SAT) Social Studies Intermediate Form II, Form W and Form X; the Elementary Devereaux Behavioral Rating Scale and the Social Studies first and fourth quarter grades as outcome variables. Specific to the SAT-Social Studies and the Social Studies grades, pre- and post-test scores were utilized. Generally, 62% of students who needed support in reading and used the cassette tapes during the intervention and the assessments (Form X) increased their post-test scores on the SAT. One third of the students were confused when presented with the presentation via paper pencil and auditorally; these students performed better when they used only paper and pencil. However, no quantitative data about their performance was noted. Mosby provided only narrative results about behavior, stating that there was an increase in comprehension, creative initiative, and a reduction in time required for work completion. Although Mosby reflected on the positive results of the study by asserting that students performed better on their SAT scores and that the face validity of the intervention was strong, the Social Studies pre-and post-test quarterly grades were not significantly different.
More recent debate about SLD models argue that the IQ-Achievement model is at best flawed. Francis, Fletcher, Stuebing, Lyon, Shaywitz, and Shaywitz, (2005) illustrated the inherent flaw of using psychometrics to identify students as learning disabled or non-disabled when decisions were made using cut-point scores. The authors’ argument was that the relationship between IQ and achievement contained measurement error, and that the relationship was statistically imperfect creating an instability in the identification process overtime.

Archival data from the Connecticut Longitudinal Study, kindergarten cohort beginning in the 1983/1984 school year was utilized. Demographic data indicated that the 403 participants from this group were 85% White, 11% Black, 2% Hispanic and 1% Asian from middle to upper middle class backgrounds. Participant data for the analysis was derived from the cohort’s 3rd and 5th grade school year. The authors chose these years as Time One and Time Two data because of previous research that reported younger students’ IQ results as more unreliable. The IQ-Achievement measures were the Wechsler Intelligence Scale for Children, Revised Full Scale and the Woodcock-Johnson Psychoeducational Test Battery Reading Composite scores, respectively. Based on a priori discussions about learning disabled and non-disabled profiles, four subgroups were determined: typically achieving, low achieving, IQ discrepant, and low achieving and IQ discrepant only.

Parallel simulated data were generated by inputting real IQ-Achievement scores from a sample of 420 students into the Statistical Analysis System (SAS) software package. Data from the 420 students provided a set of simulated or observed scores. The simulated data provided (1) typical IQ-Achievement measures that were low in
variability (i.e., highly stable and reliable correlations that were $r = .80$ or higher); and (2) hypothetical “disabled” groups that were based on a 1.5 standard deviation discrepancy between IQ and achievement and on achievement scores that fell below the twenty-fifth percentile. Once these parameters were inputted, the SAS software generated another data set allowing the authors to have simulated Time Two data.

The results from the simulated data were presented first. As predicted, all four groups (i.e., typically achieving, low achieving, IQ discrepant and low achieving, and IQ discrepant only) demonstrated change in how students were identified overtime. Specifically, the percent change across all four groups ranged between 11% and 67%. Next, the results from the CLS cohorts were presented, revealing a 9% to 32% change among the disabled and nondisabled profiles of the 3rd and 5th grade cohorts. As the authors stated, the study’s purpose was not to challenge all the limitations associated with using simulated data, but to demonstrate that the IQ-Achievement identification model was flawed and that cut-points were arbitrary. Francis et al., (2005) put it succinctly “…a single assessment at a single point in time is not psychometrically adequate for determinations that have significant long-term impact on a child’s development” (p.104).

Vellutino, Scanlon, Small, and Fanuele, (2006) demonstrated that ensuring high quality instruction reduces the risk of over identification in special education. Vellutino et al. (2006) used archival data to distinguish between cognitive and instructional deficits in a sample of 1284 students. Students were identified in the middle of first grade as normal or struggling readers. Baseline data about cognitive profiles and foundational literacy skills (e.g., letter identification, initial sound fluency) were retrieved and examined. In addition to general instruction in the first and second grade curriculum,
struggling readers were given one-to-one tutoring for two semesters (i.e., mid-first grade to mid-second grade) depending on need.

Post data analysis indicated that only 1.5% of the struggling readers were later identified as students with disabilities. Vellutino et al. (2006) reported that previous eligibility percentages could have been as high as 9%. Further analysis of the data revealed that there were some students who, with tutoring support, scored in the average ranges on basic word skill assessments, but scored in the below average range when the intervention period ended. In order to explain the behaviors of this subset in the sample, a review of these students’ skill profiles revealed differences between at-risk students and remediated students. At-risk students (1) demonstrated the least initial growth during the early stages of the intervention, (2) scored lower on the reading tests (e.g., the mean score on the Letter Identification on the Woodcock Reading Tests, Revised was 5.75, in comparison to the non-at-risk mean score of 25.6), and (3) were deficient in other emergent reading skills upon entry in kindergarten. Vellutino et al. noted another important finding about at-risk and remediated students, which was that intelligence quotient (IQ) scores were not different for the two groups of students. Vellutino et al. were asserting that the IQ-achievement discrepancy was not significantly correlated with reading growth in the early grades, but that reading growth was more linked to balanced, explicit instruction and the most appropriate reading intervention. In effect, reading growth is correlated with designing “specialized” instruction and interventions to meet the unique needs of students.

Although the Mosby’s (1979) research did not demonstrate a significant intervention effect, its empirical focus was on determining how general education
instruction coupled with a “specially designed instruction” could lead to increased academic achievement for students with learning disabilities. Current research and best practice surrounding SLD and/or at-risk students seek similar outcomes. What is different is an understanding of the multifaceted nature of SLD as a construct and, as a consequence, the approach by which educators intervene (Hale et al., 2010; National Research Center on Learning Disabilities, 2007; Vellutino, Scanlon, Small & Fanuele, 2006). In light of this shift, a revised research approach to understanding diverse students with SLD is also warranted.

To summarize, the SLD construct was originally defined as a manifestation of uneven cognitive abilities that interfered with students’ learning. Eligibility for special education was based on academic need and the severity between one or more cognitive processes and a student’s academic achievement. Research surrounding the best practice of the discrepancy model demonstrates that this process is not sufficient in meeting the needs of students with learning disabilities, in particular students with SLD from diverse backgrounds. More recent research that explores revised models of the SLD construct, including more effective use of explicit instruction and “specialized” intervention practices, demonstrates that less students are in need of special education services (Francis et al., 2005; National Research Center on Learning Disabilities, 2007; Vellutino et al., 2006).

**English Language Learners.** Recent data suggest that there are 11.2 million (or 21% of the school-aged population) ELL children speaking over 400 languages in US schools. Of this number, Hispanics (e.g., Mexicans, Puerto Ricans or Cubans) account for at least 70% of this population in the US (Aud et al., 2011). As stated in chapter one,
English Language Learner (ELL) students are generally “…students in grades K-12 whose primary or first language is not English and who have not passed their states’ English language proficiency test” (p. ii) (Haas & Huang, 2010). A more comprehensive definition was offered by Lopez (1995) who reported that ELL (formally referred to as Limited English Proficient) students typically came from homes that were culturally different, communicated in a language other than English, and demonstrated proficiency in their first language’s receptive and expressive abilities.

Relative to homogeneity concerns, Artiles, Rueda, Salazar and Higareda (2005) conducted research to examine the disproportionate representation of ELL students in 11 districts in Southern California. The purpose of their study was to determine the extent to which ELL subgroups were disproportionately represented in special education or in an inappropriate grade level. Also, the extent to which an English program and special education placement were connected to grade and SES were examined. The student population across the 11 districts ranged between 52,000 and 77,000 students. Demographic data revealed a highly diverse student body with 69% identified as Latino/Chicano, 13.5% African American, 10% White and 4.3% Asian. Within this group, 42% were considered ELL; 90% of ELL students were Latino. Relative to poverty status, greater than 75% of the students (elementary and high school) received free or reduced lunch.

Artiles et al. (2005) reported that all districts measured an ELL student’s English proficiency as the ability to demonstrate reading and language skills on a standardized assessment by scoring at the 36th percentile or better. Based on 1998/1999 data, 49% of ELL fifth grade students were still coded as ELL 1 or ELL 2; the latter designation are
those students who are limited in English and in their primary language. Special education categories reviewed in this study included intellectual disabilities, speech/language impairments and specific learning disabilities. The authors used federal and state guidelines to determine eligibility for special education programs and limited the data to those students receiving direct or related services between 21% and 60% of the school day. In effect, self contained students were excluded. Overrepresentation was defined as group membership that was 10% greater than what was expected in the school-aged population.

At the elementary level, Artiles et al. (2005) found overrepresentations for English limited (L1) only and the limited English/limited primary language (L2) groups. In fact, the results indicated that L2 students were twice as likely as White students to be placed into a learning disabled program. In contrast, it was found that English proficient students were two times more likely to be underrepresented or identified as not needing learning disability services, while students with language/speech impairments were four times more likely to be overlooked as needing services. When answering the research question about language program representation, it was found that the type of language program had an effect on overrepresentation. More specifically, elementary ELL students who were in English immersion programs were two and three times as likely to be placed in more restrictive LD services when compared to students in the modified immersion and bilingual programs, respectively. As expected, the majority of ELL students who were in special education programs (except language/speech impaired students) were from low SES backgrounds; these results were independent of students’ elementary or secondary grade level.
One of the limitations mentioned by Artiles et al. (2005) was that school level variables were not included in the analysis. However, another limitation (albeit out of the control of the investigators) of the study was that the ELL categories needed to be coded to reflect a wider range of proficiency levels. It is possible that more variability exists when the ELL data are further disaggregated.

D’Angiulli, Siegel, and Maggi, (2004) also examined the effect of literacy instruction on ELL students and their native English (L1) counterparts when considering participant’s socio-economic status. The authors’ purpose was to determine if the relationship between SES and word reading development was the same for ELL and L1 students. The second purpose of the study was to determine the impact of the literacy-intensive instruction on the developmental reading skills between the two groups. Finally, the authors asked if the literacy intensive instruction reduced the risk for disadvantaged ELL learners when compared to disadvantaged L1 students.

The study employed a quasi-longitudinal experimental design using archival data from 30 schools and 1108 students in kindergarten through fifth grade. Of the 1108 participants, 75% were native English speaking and 23% were ELL students. D’Angiulli and colleagues (2004) stated that the three to one ratio was reflective of North Vancouver districts and of the province of British Columbia’s demographic profile. The North Vancouver’s literacy intensive curriculum served as the independent variable. The curriculum included explicit instruction activities that emphasized sound-symbol relationships, six reading components (guided reading, shared reading, read/write connection, home reading, independent reading and read aloud/respond), and the ‘daily dozen’ or instruction in twelve reading modules. The twelve modules were delivered via
The dependent variables in this study were word reading achievement and socio-economic status. The word reading achievement was measured using the reading subtest (Blue Form) of the Wide Range Achievement Test-3 (WRAT-3). Participants were individually assessed once in the Fall of kindergarten and once in the Spring at every grade level beginning in kindergarten; if a participant had less than three assessments their data were not included in the analysis. It was noted that a participant’s socio-economic status was based on ten indicators that were derived from the 1991 Canadian census. The authors reported reliability estimates of $r = .85$ after conducting test-re-test analyses between the 1991 and 2001 Canadian census.

The first research question was analyzed using a socio-economic gradient. Results revealed that the relationship between SES and word reading was significant for kindergarteners who were considered ELL in two of the three categories while the relationship between SES and word reading was only significant in one of the kindergarten categories for E1 students. Data also revealed that as the word reading and SES variables were assessed overtime (as participants progressed in higher grades) the relationship was less strong. In effect, as ELL students received more reading instruction, SES was less of a factor in their academic achievement.

The second and third research questions used a growth mixture modeling equation. The authors divided SES and Reading into four respective groups. Specifically, SES was divided into quartiles with quartile one being the lowest. Percentile scores from the word reading subtest of the WRAT were sub-grouped as proficient achievers (above average performers), improvers to proficiency (average to above average performers as student move to higher grades), non-proficient achievers
(independent of grade, these students remain in the at-risk or below 40th percentiles) and non-proficient improvers (demonstrate improvement with grade, but remain below proficiency). Relative to the second research question about the impact of the literacy-intensive instruction on the development of reading skills, it was noted that overtime (i.e., from K to 5th grade) most ELL students were considered proficient achievers or improvers to proficiency. Data where ELL struggled academically were noted in the non-proficient improvers category where improvements were noted across the grade levels, but proficiency status was not yet attained. Essentially, these ELL students always obtained high average scores or were successful in working toward achieving an “average” standard on assessments. It was noted that while improvements were noted for the ELL group, data revealed that the ELL students’ trajectories remained lower than their E1 counterparts from kindergarten to third grade, but were not significantly different by fifth grade. Another cluster of ELL (3%) and L1 (25%) participants started with low scores in kindergarten, but remained in the very at-risk to at-risk categories through 5th grade. The authors suggested that other factors (e.g., the need for more intensive literacy intervention) beyond SES and general literacy instruction may have contributed to these results. The authors answered the third research question and concluded that it is possible that the North Vancouver intensive curriculum instructional program may reduce the at-risk population associated with reading failure.

Although these results indicated promise for ELL students, questions remain about the ELL group who remained at risk even though they received instruction in the district’s curriculum. Closer attention needs to be given to the SES construct, as there was no mention about how the construct was divided into quartiles. It is possible that the
parents of ELL group that remained at risk may have had more extreme cases of low SES. Similarly, there was no indication about how students’ ELL status changed overtime. In other words, what was the number of ELL students who were analyzed at the (1) initial, (2) moderately fluent or (3) proficient stages?

Another longitudinal study about ELL students was conducted by Kieffer (2011). Kieffer’s purpose was to describe the English reading growth experiences of language minority youth from kindergarten through eighth grade. Also, in similar fashion to the D’Angiulli et al. (2004) study, the secondary purpose of Kieffer’s study was to determine the effect of SES on reading achievement. Further, the intent was to add to the ELL body of research because little was known about what to expect from Language Minority (LM) students and their reading achievement overtime, especially when accounting for various levels of English proficiency among LM learners. Participant data were gathered from the Early Childhood Longitudinal Study-kindergarten cohort (ECLS-K) and stored at the National Center for Educational Statistics (NCES). A multi-stage probability sampling method was utilized to increase opportunities for a representative sampling of home, student and school characteristics. The 9-year study began with 21,409 kindergarten students; however, the final sample was 9,189 eighth graders. Of these 9,189 data files, the LM data were further divided by initial limited English proficiency (LM-iLEP), initial fluent English Language proficiency (LM-iFEP) and Native English (NE) speaker. The LM-iLEP group was later further divided because during the first four rounds of assessments, the author noted a large variability among this group. Therefore, group membership was based on six categories: (1) Native English speaker (NE), (2) Language Minority initial fluent English Proficiency (LM-iFEP), (3) Language Minority initial
limited English proficient and passing the Pre-LAS in the spring of kindergarten (LM-iLEP-QUALSK), (4) Language Minority initial limited English proficient and passing the Pre-LAS in the fall of first grade (LM-iLEP-QUALF1), (5) Language Minority initial limited English proficient and passing the Pre-LAS in the spring of first grade (LM-iLEP-QUALS1), (6) Language Minority initial limited English proficient and not passing the Pre-LAS by the spring of first grade, but taking the reading assessment in the spring of third grade (LM-iLEP-QUALS3).

Reading achievement was determined by an English Reading achievement test that used grade appropriate items (e.g., phonics, vocabulary, comprehension skills) that were combined from NAEP standardized assessments and questions from previous versions of the ECLS studies. Technical properties were reported as moderate to high. The second and third outcome measures were the SES variables. The first variable was based on interview questions (e.g., questions related to parent education, income, and occupation) asked by ECLS personnel in year 1, 2, 4, 6 and 9. Because of high inter-rater reliability and collinearity of the questionnaire, a single-time latent composite was derived for this variable. The author also reported that a school related measure (i.e., based on the school’s free and reduced lunch data) was completed in the same years that the child SES variable was assessed. Also, in similar fashion, a single-time latent composite was derived.

Piecewise latent growth modeling was used to examine the NE and LM slopes, taking into consideration the child and school SES measures on reading achievement. A growth model was also fitted to explain latent growth for grades K through 1st, 1st through 3rd, and 3rd grade through 8th grade. Maximum likelihood estimation was employed to
facilitate the greatest probability of the slope estimates given that there were missing data in the sample.

When growth estimates were grouped, the data indicated that all students from the LM group had initial status’ that were below native English speakers. Moreover, it was noted that the greatest growth occurred between kindergarten and first grade; this growth was twice as large as the growth between first and third grade which was, in turn, three times as large as the growth between third and eighth grade. The LM-iFEP students demonstrated comparable achievement results to the Native English students by first grade and maintained those comparable slope trajectories overtime. In contrast, the LM-iLEP students as a group did not demonstrate proficiency levels that were commensurate with their NE peers until eighth grade, but their rate of progress was better than those students from the NE group. The author reported this finding as “substantial across time” (p. 1209). Similar results about rate of growth were noted when the author controlled for SES backgrounds. Therefore, LM students made more rapid gains in the curriculum than their NE peers overtime. Other noteworthy findings after controlling for SES was that the LM students who were not proficient in English by the spring of first grade often lagged behind (in some places two years behind) in reading achievement at least until their middle school years. Implications raised for the education field were that early intervention measures needed to be differentiated in elementary school, while reading interventions in middle and high school could look similar to those of native English speakers.

Kieffer (2011) highlighted that the one time score in reading achievement was a challenge to this study. He further added that additional longitudinal research examining
different components of the reading process (e.g., fluency, vocabulary, comprehension) may provide additional information about this population. In addition to the limitation noted in the study, it may be important for researchers to use analyses that allow for further disaggregation of constructs (e.g., SES or ELL), as it may provide the diverse population field with a deeper understanding of specific timeframes in which to intervene and raise student achievement. It is recommended that future research on Native English Speakers report their race/ethnic profile, because representation that is different from the national norms may influence results.

When taken together, research involving the ELL student population suggests that ELLs are at greater risk for academic failure when compared to non-ELL students. In particular, studies for elementary and secondary students report that ELL students are more likely to be placed in special education programs and conversely, when demonstrating a need for special services, may not be given those services. Moreover, ELL students appear to fare better when exposed to explicit reading instruction, as evidenced by commensurate standardized assessment scores between ELL students and their non-ELL counterparts. These patterns emerge as early as first grade and are also noted in middle school. The ELL literature calls for research that expands the ELL/non-ELL categories to include more proficiency levels in order to better identify and target more salient reading interventions. Also, there are suggestions in the research for more studies to examine the impact of the type of reading (i.e., vocabulary, fluency or comprehension) on ELL students overall reading achievement. Finally, further analyses that include SES as a variable should be considered, as there is evidence that SES
becomes less of a factor when ELL students obtain greater levels of instruction in reading.

**Context for Authentic Measures**

When the National Commission on Excellence in Education recommended rigorous measurable standards in its Nation at Risk report, it was not known at that time whether annual assessments would be sufficient in providing data about schools and student performance. In light of more recent findings in the literature, it is known that annual “single bar” approaches are not sufficient to hold states, schools, and students accountable (Duncan, 2012; U.S. Department of Education, 2010a; Schleicher, 2010; West & Peterson, 2003). Rather, formative or on-going evaluation approaches are needed to facilitate the easy implementation and availability of longitudinal data that inform instruction and increase the probability of success for all learners.

From a historical perspective, authentic measures or formative evaluation have roots in behavior change. Behavior change is simply replacing an undesired behavior with a desired one. It is more likely to occur when the contingencies (e.g., rewards or consequences) are external, discrete and unambiguous, and when target behaviors are valued (Cone, 1999; Kazdin, 1980). The literature also demonstrates that behavior monitoring may occur in isolation, with a group, or in multiple settings (Cole, Marder, & McCann, 2000; Korotitsch & Nelson-Gray, 1999).

**The Dynamic Indicators of Basic Early Literacy Skills (DIBELS).** DIBELS measures are technically adequate, easy to administer, less expensive than other standardized measures (including statewide assessments), fit well into a problem-solving model, are sensitive to change, and are linked to instruction (Good, Gruba, & Kaminski,
The DIBELS serve as an assessment tool for identifying and monitoring early reading and literacy skills for at-risk students within the context of a problem-solving model referred to as the Outcomes-Driven Model (ODM) (Kaminski, Cummings, Powell-Smith & Good, 2007).

DIBELS measures were developed to (a) identify students who are at-risk for reading failure early in their academic experience, (b) provide teachers with information about how to remediate learning gaps, and (c) to answer questions about how well students are making progress toward specific reading skill area (e.g., phonemic awareness, fluency with connected text) goals (Good et al. 2001; Kaminski et al., 2008). Repeated findings in the reading literature indicate that early identification and remediation (using research-based interventions) of reading problems reduces the achievement gap and increases the likelihood that students are successful readers at the fourth grade level (National Institute of Child Health and Human Development, 2000; Stanovich, 1986).

In addition to its use as a screening and diagnostic tool, DIBELS measures also may be used within the outcomes-driven model to increase an individual’s, or a school’s reading goal. For example, the measures may be used to monitor student progress, to compare or evaluate the efficacy of instruction or supplemental support, and to determine if students are meeting reading goals and expectations. In order to accomplish this, the developers of the DIBELS created minimum quarterly benchmarks so that educators could determine the probability of reading success or failure. Specifically, when assessed using DIBELS oral reading fluency passages, students’ scores provide educators with information about the intensity of instructional support that may be needed in order for
students to be successful. For example, a strategic rating suggests that a student’s chance of meeting his next benchmark is at or less than fifty percent. Similarly, outcome data from these ratings (i.e., low risk, strategic and intensive) may also guide supervising teachers, teams, and district personnel on how to help a school develop tiered levels of instructional support and or supervision (Kaminski et al., 2008).

Four specific measures (i.e., Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency and Nonsense Word Fluency) were part of the initial DIBELS’ development and norming process between 1997 and 2001. The authors noted that the DIBELS oral reading fluency (DORF) was gathered from the Test of Oral reading Fluency (TORF), as developed by the Children’s Educational Services in 1987. Retell Fluency (RTF) and Word Use Fluency (WUF) were later added to increase the scope of the measures. The purpose of RTF (Good, Kaminski, Dill, 2002) is to (a) prevent learning or practicing a misrule (i.e., speed-reading without attention to meaning), (b) identify children whose comprehension is not consistent with their fluency, (c) provide an explicit linkage to the core components of the National Reading Panel report, and (d) to increase the face validity of DORF. WUF provides an index of vocabulary use and oral language development. In all, the DIBELS consists of seven measures that serve as indicators of the 5 Big Idea areas of reading development: phonemic awareness, phonics, vocabulary, comprehension and fluency. However, because the DORF is the only measure of interest to this study, only its technical adequacy will be reported in the following section.

As mentioned above, the DORF passages and administration procedures were initially gathered from the TORF passages. Research on the technical adequacy of the
TORF suggests moderate to strong relationships for validity (i.e., $r = .52$ to $r = .91$) and reliability co-efficients (i.e., $r = .92$ to $r = .97$), respectively (Good & Jefferson, 1998; Tindal, Marston & Deno, 1983). Also, research demonstrated no significant differences between TORF probes and other curriculum-based probes when monitoring students’ oral reading fluency over time. In other words, TORF reading probes appear to be as effective as curriculum-based measures from basal texts in assessing students’ reading progress and in assisting with critical decision making (Powell-Smith & Bradley-Klug, 2001).

The literature on DIBELS overwhelmingly support its technical adequacy as an authentic measure. Further, they demonstrate their effectiveness as effective screening and monitoring measures, especially in reading. Essentially, with over 30 years of research, the DIBELS are researched based tools that can determine growth and inform instruction (Fuchs, 2004).

**The DIBELS and State Assessments.** Research demonstrates that the DIBELS is technically strong (i.e., in reliability and validity) and is considered a very good predictive, yet inexpensive assessment tool in many school districts within the US (Buck & Torgesen, 2002; Good, Simmons & Kame’enui, 2001). Good et al. (2001) conducted a longitudinal study examining the utility and predictive power of the DIBELS’ measures with reading outcomes on the Oregon Statewide Assessment (OSA). In addition, their study supported the linkages across measures for decision-making purposes. Participants totaled 378 students from four cohorts across six elementary schools. Participants were followed from kindergarten through third grade. The participants were part of an urban school district in the Pacific Northwest where as much as 63% of the students qualified
for free and/or reduced lunch services. The authors reported that 10% and 18% of the students were considered minority and below the poverty range, respectively.

Two independent measures were used: the DIBELS fluency measures and TORF fluency measures. These two measures were used to predict scores on the OSA-Reading/Literature measure. The results of this study revealed a positive relationship between the TORF reading measure and the OSA ($r = .67$). Specifically, Good et al. found that 96% of the students who met the third grade benchmark goals (i.e., 110 words read correct per minute or more) met or exceeded the minimum requirements on the OSA. In contrast, only 28% of those students who read less than 70 words met expectation on the OSA.

When discussing the implications, Good et al. (2001) stated that the benchmarks had strong utility, especially in the lower grades where targeted remedial instruction would be more useful. In fact, they reported that ignoring the low performance of students relative to foundational reading processes jeopardized the “high stakes” outcomes. Good et al. demonstrated that longitudinal studies help us better understand research. In this case, the longitudinal focus of the study is analogous to ongoing progress monitoring on a quarterly basis. Further, they adequately demonstrated why using authentic measures was important to research, in that they might be considered “a compliment” to statewide assessment.

One shortcoming of the research was that the results were not generalizable beyond the reported district. Good et al. (2001) noted this and also reported that a comparative study between districts would provide insight into ineffective and effective
instruction methods. Also, this study might have found a stronger correlational
relationship if the researchers used ongoing monitoring.

Similarly, Shaw and Shaw (2002) examined the concurrent predictive validity of
the DIBELS and the Colorado State Assessment Program. A total of fifty-two third
grade students from a Colorado elementary school were administered three assessments
of the DIBELS (i.e., in the fall, winter and spring) and also the spring version of the
Colorado State Assessment Program (CSAP). Participants also included teachers and
Reading Center Staff who were trained in DIBELS’ administrative procedures just prior
to the first administration in the fall 2001.

The spring benchmarks for third graders were 110 words read per minute and a
proficient or advanced level on the DIBELS and CSAP, respectively. A proficient level
on the CSAP meant that a student earned 526 or more points. More definitively, a
proficient level meant that a student demonstrated adequate comprehension skills that
included making inferences, identifying the main idea with supporting details,
sequencing, drawing conclusions, and determining cause and effect. Shaw and Shaw
(2002) explained further that the level system was linked directly to the Colorado Model
Content Standards so that items on the CSAP reflected the state standards. In addition to
reporting on the face and content validity of the CSAP, Shaw and Shaw reported
adequate reliability coefficients as well.

The results revealed a strong correlation between the DIBELS and the CSAP.
That is, the Spring DIBELS and the Spring CSAP yielded a coefficient of \( r = .80 \). The
fall and winter correlations were similar (i.e., \( r = .73 \)) to the spring results. Further
analysis of the data revealed that a cut score of 90 words read correct meant that 91% of third graders were considered proficient or advanced on the CASP’s reading assessment.

Like the Good et al. (2001) study, this study adequately demonstrated how well the DIBELS complimented a statewide assessment test. Additionally, the results reinforced the utility of the DIBELS as an ongoing authentic assessment tool. However, the Shaw and Shaw (2002) study raised several questions about internal and external validity. First, they did not report any data relative to the demographics of the student participants. Finally, the Shaw and Shaw study also used a small sample that consisted of only one school.

Similarly, a study by Barger (2003) also had a small sample (i.e., 38) of third grade participants from one school in North Carolina to determine the relationship between the DIBELS ORF (spring) and the North Carolina End of Grade Test. The two assessments were administered one week a part. The results of the correlational analysis showed a positive correlation (i.e., $r = .73$) between the two measures. Even though the analysis was one-week a part, Barger posited that the DIBELS ORF might be an accurate predictor of grade level proficiency on the North Carolina End of Grade Test reading assessment. More descriptively, the higher the score on the DIBELS, the more likely it was for a student to obtain a level three or higher on the North Carolina End of Grade test. Barger reported that 100 words read correct per minute seemed to be an appropriate cut score for making accurate decisions. Conversely, Barger’s results also stated ORF scores between 70 and 99 were much harder to predict, as only 50% of the students passed the North Carolina End of Grade test.
In addition to increasing sample size, a replication of the results needs to focus on internal reliability (e.g., data collector training and inter-rater reliability) and generalizability. Further, like the other studies, Barger’s (2003) study provided support for using DIBELS to predict outcomes on a statewide assessment; however, it does not provide critical information about the level or type of DIBELS assessment monitoring that is most effective when predicting high stakes assessment.

Studies using the DIBELS as a complementary measure were also conducted in the State of Florida. Specifically, Buck and Torgesen (2003) examined the relationship between the DIBELS oral reading fluency (DORF) and the Florida Comprehensive Assessment Test (FCAT) to determine if the DORF measures were good predictors of reading comprehension.

The study consisted of 1102 (49% were female) participants who were chosen from one of the 67 districts in Florida. Additional demographic indicators revealed that 83, 7, and 6 percent of the participants were White, Black, and Hispanic, respectively. Forty-Six percent of the students received free or reduced lunch, and 19% received exceptional student education services. The predictor variable was the median DORF score and the criterion variable was standard scaled scores on the FCAT (Sunshine State Standards-SSS).

The results of the study revealed a significant correlation between the DORF and FCAT-SSS scores ($r = .74$, $p < .001$). It was also noted that similar correlations were found between the DORF scores and the FCAT-Norm Referenced Test (NRT) scores ($r = .74$, $p < .001$), and between the math and the DORF scores ($r = .53$, $p < .001$). When additional correlation analyses were applied based on DORF reading categories (i.e., high
medium or low risk) they found that students who read 110 words or more correct per minute were considered low risk. Stated differently, these students (91% in the sample) were more likely to score at a proficient level (i.e., level 3 or better) on the FCAT. Conversely, only 19% of the students reading less than 80 words correct per minute reached a level 3 or better on the FCAT. When Buck and Torgesen (2003) analyzed the data by subgroup, they found that attaining 110 words correct per minute did not correlate as well for minority students. In other words, having a high DORF score was less of a predictor of proficient reading performance on the FCAT for disadvantaged students. The authors hypothesized that minorities (especially African Americans) have other reading deficits (e.g., vocabulary) that impede their success. Buck and Torgesen found that reading less than 80 words read correct per minute was a strong predictor of failure on the FCAT for minority students.

One of the advantages of this study was the large sample size. However, a major limitation to the study was the small representation of ethnic minorities. A replication of this study or a partial replication that includes more minority groups is warranted, especially given that minority students, across the nation, are the most challenged in making the grade on high stakes state assessments, inclusive of the FCAT (Thernstrom & Thernstrom, 2003; U.S. Department of Education, 2011a).

Similarly, Cook (2003) examined the relationship between the DIBELS and the Stanford Achievement Test, Ninth Edition (SAT-9). Her specific research question was: What is the concurrent validity of the DIBELS and the SAT-9? Participant data were gathered from archival data on five 1st grade classrooms of a school in Ohio. Specifically, the reading data of 79 Caucasian students (40 females and 39 males) from general and
special education classrooms participated. Fifty-seven of these students received free and reduced lunch.

Specific to the measures, Cook (2003) reported that the SAT-9 was administered in the spring using the group standardized administration directions. Within the SAT-9 Cluster, the Total, Word Study, Word Reading and Reading Comprehension raw scores were obtained. Specific to the DIBELS, the Nonsense Word Fluency (NWF), Phonemic Segmentation Fluency (PSF), and DIBELS Oral Reading Fluency (DORF) measures were administered in accordance with the DIBELS standardized procedures. However, the author noted that sixteen PSF scores were missing. Also, it was reported that some students were absent, so the numerical information between students was unequal. Pearson Product Moment Correlation analyses were used to determine the correlation results between all components of the SAT-9 and the DIBELS.

The results indicated that there was a positive relationship on all PSF and SAT-9 variables except for Word Reading ($r = .179, p = .161$). The NWF was significant for all of the SAT-9 subtest areas ($r = .571, p = .000$ to $r = .639, p = .000$). Also, Cook found that the DORF and the SAT-9 scores had moderate to strong correlation ($r = .610, p = .000$ to $r = .749, p = .000$).

Consistent with Buck and Torgesen (2003), Cook (2003) reported that the strong correlation between a standardized criterion measure and the DIBELS suggest that the DIBELS may be used as an effective diagnostic and predictive instrument for schools and teachers. Also, Cook highlighted the two major limitations of her study: the imbalance in minority representation and the fact that the missing data may be the reason why the PSF results were not significant. Although including a more representative sample of
minority students in a replication study is a good idea, disaggregating data among diverse student groups is also critical because minority students are often left behind.

In addition to using grade level benchmarks to predict performance on high stakes tests, Wilson, (2005) conducted a study to also determine if the relationship between the DIBELS and the Arizona Instrument to Measure Standards (AIMS) was similar by subgroup (i.e., by gender, race, or socio-economic status). Two hundred and forty-one third graders across three schools in one of Arizona school districts participated in the study. The author noted that all three schools were considered Reading First schools. Consistent with other studies, Wilson used the DIBELS as the independent or predictor variable and reported third grade benchmarks that were consistent with the DIBELS‘ test developers. Specifically, reading less than 80 words per minute fell in the “at risk” category, reading between 80 and 109 words per minute fell in the “some risk” category and reading 110 or more words per minute fell in the “low risk” category. The dependent variable was the AIMS. The AIMS is a multiple choice standardized test that is designed to measure grade level proficiency. Scale scores and proficiency levels were used based on the spring assessment of the test.

The results from this study indicated that there was a moderately positive correlation ($r = .741$) between the DIBELS and the reading comprehension portion of the AIMS standardized assessment. When analyzed by proficiency level, Wilson (2005) reported that 81% of students in the “low’ risk category were proficient, while only 51% of students in the “some risk” were proficient. Essentially, students reading 110 words or more per minute had a higher than chance probability of reaching grade level proficiency on the AIMS test.
Wilson’s (2005) subgroup analysis indicated that being White and/or female increased the chances of being proficient on the AIMS assessment. Nonetheless, it was noted that a significant positive relationship existed (a range between $r = .669$ to $r = .781$) between the DIBELS and the AIMS for minority groups. Essentially, the results were similar for demographic subgroups as well.

The consistency of the results in this recent study, when compared to prior research involving the DIBELS and other state assessment is noteworthy. However, like Wilson (2005) reported, a broader sample is needed to replicate results. A broader sample is necessary to encourage a representative sampling of the demographic subgroups that comprise the US population. In this study, Wilson did not provide demographic statistics for all the major race/ethnic groups. For example, there was a category for non-English language learners (Non-ELL) for not for Blacks, Bi-racial or Asian subgroups. Finally, although the students in this study were monitored quarterly, as the schools were supported by a Reading First grant, Wilson did not determine if ongoing monitoring of students’ reading impacted proficiency levels on the AIMS.

Educators in Ohio districts are using the DIBELS as a complimentary benchmark measure as well. Van Meer, Lentz, and Stollar (2005) conducted a study to determine the relationship between the DIBELS and the reading portion of the fourth grade Ohio Proficiency Test (OPT). The participants of this study were 364 students from three schools in a suburban district of Ohio. Participants’ data were gathered when students were in the third and fourth grades, respectively. The demographic revealed that the majority of the student participants were White; less than 25% of the students were considered economically disadvantaged. Also, students with individualized education
plans were included in the study and were considered to not have “significant cognitive disabilities.”

The two independent measures were the CBM ORF and the DORF. The dependent variable was the Ohio Fourth Grade Reading Proficiency Test (OPT). The CBM ORF was administered in the fall and spring, and the DORF passages were administered to the participants during the fall, winter and spring of 3rd grade. Benchmark goals used to identify at-risk students were taken from DIBELS benchmarks and CBM 4th grade data. Like other state assessments, the reading portion of the OPT was administered to all fourth graders in October. Students were given additional opportunities to pass this test in March and July. The thirty OPT items are based on the fourth grade learning outcomes that examined fiction and nonfiction texts. The response format included multiple choice, short-answer, and extended response.

The results supported moderate, but significant correlations between ORF and OPT for reading. More specifically, the correlations ranged from $r = .61$ to $r = .65 (p = .01)$, and were judged to be of similar pattern across the fall, winter and spring of the students’ 3rd and 4th grade years. In both years, at least 75% of the students passed at the proficiency level on the OPT when the minimum benchmark goals ranged between 93 to 110 words read correct per minute. The authors also noted that most of the 4th grade students who were identified as at risk during the fall administration were most unlikely to be successful on the OPT, despite the availability for two additional attempts. The highest false positive for this group was 26%. The error was attributed to the efficacy of the reading interventions between the fall and spring administration of the OPT. Van
Meer et al. (2005) also found that the DIBELS and CBM criteria were valid for end of year benchmarks and thus, were appropriate to use for individual goal setting.

Although the results of this study were consistent with previous prediction studies involving DIBELS and statewide assessments, Van Meer et al. (2005) noted that the predictive relationship was less strong when compared to other studies (e.g., Barger, 2003; Buck & Torgesen, 2002). They explained that the differences might have been due to the fact that constructs (e.g., critical thinking) in addition to reading were being measured. The authors also offered that using the fall scores to identify at risk students for interventions, as well as to predict the OPT in the fourth grade was a limitation to this study. Also, there was indication that the three administrations of the OPT were conducted using alternate forms. If the same instrument was given, this was a limitation that rendered the results invalid. Finally, increasing the number of schools and the student sample size may provide more precision in the correlation regression. The implications for future research, instruction and service allocation are crucial.

In summary, research with DIBELS has demonstrated moderate to strong reliable relationships with standardized and criterion related assessments. A re-inspection of the literature review above indicates that researchers have often included diverse students (e.g., Blacks, rural, urban, and students with disabilities) as subgroups in their samples; however the diverse samples were often small and not the main focus of the research. Of those studies that had reasonable sample sizes about diverse students, the results have demonstrated that diverse students are at greater risk for academic failure (Buck & Torgesen, 2002; Good et al, 2001; Wilson, 2005). The next section of this chapter
examines recent research examples of how the DIBELS measures fare with diverse students.

**DIBELS Measures and Diverse Students**

A recent study by Baker et al., (2008) is an example of research that examines data about diverse students indirectly based on its large sample of diverse students. In the Baker et al.’s case, 69% of the students were low SES and 32% qualified for ELL services. The purpose of their study was three fold. First, the authors wanted to determine how well ORF correlated with the Oregon Statewide Reading Assessment (OSRA). Second, they wanted to examine the unique ORF slope contributions when predicting reading comprehension outcomes, while controlling for initial levels of reading performance. Finally, the study’s purpose was also to examine various statistical models in order to determine how well ORF would predict comprehension performance from year one to year two.

Data files of 9,600 students from 34 Reading First schools were included in the study. Fifty percent of the schools from the sixteen districts were considered urban, while the remaining schools were considered rural to midsize. As mentioned earlier, 69% of the students in the district qualified for free and reduced lunch and 32% qualified for ELL services; 10% of the students qualified for ESE services. Data for the study were collected during the first two years of Reading First implementation in the State of Oregon. The authors divided the student data into four cohorts: cohort one comprised 3rd grade student data in year one of data collection; cohort two comprised 2nd and 3rd grade student data in years one and two; cohort three comprised 1st and 2nd grade student data collected in years one and two; and cohort four comprised 1st grade data collected in year
two. Data checks revealed that 10% to 13% of the second and third grade data were missing, while 5% to 7% at the first grade level had missing data. Students with missing data were included in the study as long as they had one predictor score (e.g., DIBELS ORF) and a score on the outcome measure because the type of analysis used was robust to missing data.

The predictor variables chosen for this study were the DIBELS ORF and the Stanford Achievement Test, Tenth Edition (SAT-10). Specific to the DIBELS, Baker et al. (2008) reported that the DIBELS measures were administered to students during three assessment windows in Oregon (i.e., in the Fall, Winter and Spring). The standard administration protocols that were established by the DIBELS developers were prescribed. Similarly, the authors reported that the SAT-10 was administered based on the developers’ guidelines for first through third grade students. The dependent variable of the study was the Oregon Statewide Reading Assessment (OSRA). It is an untimed, multiple choice reading assessment that is administered to third grade students. Students are required to read literary, practical or informative selections and then are required to answer literal and inferential type questions. Strong criterion related validity was established with measures like the California Achievement Test and the Iowa Tests of Basic Skills.

Descriptive results of the four cohorts indicated that the range between groups was not significant. Relative to the first research question, correlation regressions revealed moderate to strong positive relationships between the ORF, the ORSA, the SAT-10 first grade, and the SAT-10 second grade (i.e., a range between $r = .58$ and $r = .80$). Slope contributions to reading comprehension were answered by examining data from
second graders (i.e., cohorts two through four), and third graders (i.e., cohorts one through three). The predictors employed in the second grade analysis were (1) ORF intercept, (2) ORF intercept and slope, (3) ORF intercept, slope and SAT-10 (first grade assessment), (4) SAT-10 (first grade assessment), (5) ORF slope and SAT-10 (first grade assessment), and (6) ORF intercept and SAT-10 (first grade assessment). Results revealed that together the intercepts and slopes accounted for 70% of the variance on the second grade SAT-10 (p < .0001) and that the ORF slopes accounted for an additional 10% of the unique variance when controlling for initial levels of performance. Similar predictors were used in the third grade analysis. The intercepts and slopes accounted for 52% of the variance on the OSRA assessment; an additional 3% unique variance was credited to the 3rd grade slope alone.

Relative to the third research question (i.e., predicting comprehension performance from year one to year two) data from the first and second grades revealed that the first grade SAT-10 was the strongest predictor, followed by the ORF slopes and then the ORF intercept. Together, these predictors accounted for 76% of the variance. Data from the second and third grade revealed that the ORF intercept explained a greater portion of the variance over and above the ORF slopes when the outcome measure was the OSRA.

The authors summarized their findings by stating that their correlation findings extended the research regarding the ORF as a valid complimentary measure to high stakes and commercially standardized reading measures. Also, they noted that the unique contributions of the reading slopes (when controlling for the intercept) was supported at the second and third grade level, though its unique contribution declined from 10% to
3%; both slopes variances were statistically significant. It is also noteworthy that the results appear to support prior research which found that the ORF alone may not be the best predictor for high risk populations (Buck & Torgesen, 2002). In this study’s case, reading comprehension (i.e., first grade SAT-10 scores) accounted for more of the variance over and above ORF scores for second graders. Finally, the Baker et al. (2008) study’s large sample size increased the external validity of the research. However, a limitation was that no analyses were conducted with sub samples of diverse student groups to explore the impact of the intercepts and slopes. The authors noted that this was a limitation of their study.

Millet (2011) chose to examine the ability of the DORF measures to predict comprehension for ELL students. In particular, Millet wanted to investigate the correlations between first through third grade DORF measures and the TorraNova Reading assessment measures. Also, he was curious about whether the Peabody Picture Vocabulary Test, third edition would further explain any of the unexplained variance between the DORF and TorraNova Reading variables.

Participant archival data was retrieved on 65 Hispanic students in Arizona who qualified for free or reduced lunch. The students were assessed upon entry to kindergarten with the Arizona English Language Learner Assessment (AZELLA) and scored at the emergent range; the students did not have any prior pre-school experiences. Like other state language assessments, the AZELLA has cut scores to determine when a student is considered English proficient. A proficient level is the highest and the pre-emergent level is the lowest. No school level characteristics were mentioned except that
all the assessment scores used in the analysis were administered in the Winter and Spring of the students’ respective 1st, 2nd and 3rd grade years.

Descriptive analyses suggested a positive relationship existed between the 1st, 2nd and 3rd DIBELS scores and the 2nd and 3rd grade TerraNova Reading comprehensive assessment test, respectively. Inferential statistics revealed that all correlations were statistically significant ($p < .05$); however, it was noted that the first grade correlation scores were higher than the second grade. As expected, the correlation between 3rd grade DIBELS and the 3rd grade TerraReading ($r = .68$) was the highest. Multiple regression analyses were conducted in order to determine how much of the vocabulary and reading fluency variances explained reading comprehension for this group of ELL students. Millett (2011) found that the vocabulary measure explained about half of the variance (e.g., 47% of the variance) in the earlier grades; however, an inverse relationship was noted when third grade DIBELS scores were correlated with the TerraNova Reading criterion and the PPVT-III, respectively. In effect, reading fluency explained 45% of the variance. Millett surmised that the results in his study supported previous research that found that vocabulary knowledge was a better predictor of reading comprehension in the early grades. More importantly, Millett’s research reiterated that the educators should be careful when using the DIBELS assessments as a sole measure with all students, but especially young ELL students. This caution was offered due to the fact that the oral fluency mean in 1st grade exceeded the benchmark expectation, yet only accounted for 7% of the explained variance.

Although the overall results suggested a positive relationship between the DIBELS and the Arizona state assessment, the researcher noted the need for future
research among ELL students. In particular, Millett suggested that other studies might extend the geographical area to include increased sample size and a greater diversity of ELL students. Also, he suggested that researchers disaggregate the data within the ELL sample to determine if there are differences based on level of ELL classification and or socio-economic status.

**Summary**

Despite years of education reform, America’s efforts toward closing the achievement gap between diverse learners and their receptive counterparts has yet to be attained (Aud et al., 2011; National Center of Education Statistics, 2011). More specifically, empirical data about students with learning disabilities and ELL students suggest that these two diverse groups continue to be at greater risk for academic failure. Research in the last decade demonstrates the DIBELS oral reading fluency measure provides moderate to strong predictive validity with high stakes state assessments, and is relevant for identifying skill gaps, informing instruction, and monitoring progress for all students including diverse students (Baker et al., 2008; Buck & Torgesen, 2002; McGlinchey & Hixon, 2004; Stage & Jacobsen, 2001; and Wilson, 2005). However, one consistent limitation noted in the DIBELS literature is that there are few longitudinal studies that have purposefully investigated the impact of authentic measures on diverse student subgroups. In essence, investigations about the unique contributions of diverse students’ initial reading levels and their reading slopes are warranted. Moreover, many of the studies reviewed suggest that future longitudinal analyses consider increasing school and student sample sizes so that results investigating intra-group reading differences are valid.
Chapter Three

Method

Overview

The purpose of this study was to examine the intra- and inter-group growth trajectories of two diverse student groups on reading achievement. The study employed a longitudinal, quasi-experimental research design to answer the research questions. The first research question examined the extent to which the intercepts and slopes of third grade reading trajectories differed between English Language Learners (ELLs) enrolled in an ESOL program (LY), ELLs not enrolled but monitored (LF), ELLs not enrolled and no longer monitored (LZ), and non-ELLs when holding gender and SES constant. The second research question investigated the extent to which the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differed across racial/ethnic groups. The chapter begins by identifying the total number of participant data and the number of observations or time points utilized in both research questions. Second, frequency data are presented on the independent and dependent variables. Third, information about how the independent and dependent variables are defined, operationalized and coded is outlined. The fourth section of this chapter describes the study procedures. Specifically, procedures for training, administration, and data collection of the DIBELS oral fluency measure (DORF) are reported. Fifth, procedures related to data retrieval and the data check processes are specified. The
chapter concludes by specifying the study design and the data analyses employed to answer both research questions.

**Participants**

**Initial Student Data Retrieved.** Three districts (i.e., two in the West/Central region and the other in the South East region) participated in the present study. Of these three districts, one was part of the *Reading First* initiative (District A) while the other two were considered Non-*Reading First* districts (Districts B and C). *Reading First* districts were districts that were provided funding from the federal government’s *Reading First* grant program through sub-grants awarded by state education agencies. The grant award was the major difference between a *Reading First* and Non-*Reading First* district or school, as Non-*Reading First* districts and schools were provided with similar training and resources, a point that is discussed later in this chapter. The main purpose of federal *Reading First* grants was to support scientifically-based reading practices (e.g., ongoing professional development for reading teachers, use of data to support instructional grouping, and use of assessment measures to monitor student progress) that facilitated the NCLB’s mission of all students reading on grade level by the end of third grade (Gamse, Jacob, Horst, Boulay, & Unlu, 2008).

The total data set retrieved from the three districts comprised student data from 28,259 third graders. The sample reflected data that were collected over three consecutive years (i.e., 2006, 2007, and 2008) in each district. Descriptive analyses of the data across districts revealed that the average number of males was 52% and the average number of females equaled 48%. Table 1 summarizes the data based on race/ethnic membership between White, Black and Hispanic students in the three
districts. A comparison of Florida third grade students’ overall race/ethnic demographics for 2006 also is included.

Table 1

<table>
<thead>
<tr>
<th>Study Sample: Third Grade Race/Ethnicity Membership by District</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Membership</td>
</tr>
<tr>
<td>District A</td>
</tr>
<tr>
<td>District B</td>
</tr>
<tr>
<td>District C</td>
</tr>
<tr>
<td>Florida (2006)</td>
</tr>
</tbody>
</table>

Note: The 2006 3rd grade Florida race/ethnicity percentages were obtained from the Education Information and Accountability Services, Florida Department of Education.

**English Language Learners.** The ELL data set (i.e., ELL and Non-ELL students) was comprised of 26,967 third grade students from 281 schools located in two districts (i.e., one in South East Florida [District A] and the other in West Central Florida [District C]). Of this number, 20,025 ELL and Non-ELL student files, along with their respective 61,248 time points (i.e., data collection measures for fall, winter, and spring) were employed in the data analysis. ELL data from District B were not available because the initial data request did not specify a request for ELL student data and the data support personnel in District B were not available to retrieve ELL data when a second request was made.

As mentioned in chapter one, English for Speakers of other Languages (ESOL) or ELL members represent a heterogeneous group consisting of many nationalities and languages. In order to provide an approximation about how many languages were represented in the present data, a list of the top five languages in District A and C (retrieved from the Florida Department of Education, Office of Academic Achievement through Language Acquisition in 2006) were presented in Table 2. In the State of Florida
and within this data sample, ELLs were represented in four main categories: ELL-LY, ELL-LF, ELL-LP, and ELL-LZ. The ELL-LY category represented students who were eligible and enrolled in programs and or services designed for ELL students. The ELL-LF category consisted of students who were monitored for two years after exiting an ELL program. ELL-LZ represented that category of students who exited an ELL program and who were no longer monitored. Finally, the ELL-LP consisted of students who were pending a language assessment based on responses to a home-language survey (given to all students and parents upon entrance in a Florida public school) or students who were considered English proficient but needed a supporting reading or writing assessment to be categorized as ineligible for ESOL services. The number of students in this last category was small, thus they were excluded from the data set. Non-ELLS were students who were determined to be Native English speakers (based on the home language survey) and therefore were not eligible for ELL programs or services. Non-ELLS were not categorized as ELLs, but for data analyses purposes they were referred to as a comparison group to ELL students. A comprehensive discussion about how Florida schools define and determine eligibility and ineligibility for ELLs and Non-ELLS follows. Table 3 provides a descriptive summary of the ELL and Non-ELL students in the study database.
Table 2

Native Languages represented among ELLs in Districts A and C in 2006

<table>
<thead>
<tr>
<th>Group Membership</th>
<th>Native Language</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>District A</td>
<td>Spanish</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>Haitian-French Creole</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Portuguese</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>Chinese-Zhongwen</td>
<td>1%</td>
</tr>
<tr>
<td>District C</td>
<td>Spanish</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>Vietnamese</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>Russian</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Serbo-Croatian</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Arabic</td>
<td>2%</td>
</tr>
</tbody>
</table>

Table 3

Demographic Data for the ELL and Non-ELL Students in the Data Set

<table>
<thead>
<tr>
<th>Group Membership</th>
<th>Gender Male</th>
<th>Gender Female</th>
<th>ELL-LY</th>
<th>ELL-LF</th>
<th>ELL-LZ</th>
<th>Non-ELLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>District A</td>
<td>52.1%</td>
<td>47.9%</td>
<td>10.6%</td>
<td>6.1%</td>
<td>5.8%</td>
<td>77.6%</td>
</tr>
<tr>
<td>District C</td>
<td>53.1%</td>
<td>46.9%</td>
<td>5.5%</td>
<td>2.5%</td>
<td>3.1%</td>
<td>88.8%</td>
</tr>
<tr>
<td>Total</td>
<td>52.3%</td>
<td>47.7%</td>
<td>9.7%</td>
<td>5.4%</td>
<td>5.3%</td>
<td>79.6%</td>
</tr>
</tbody>
</table>

Note: The total number of the ELL/Non-ELL sample was 26,967. Of this total, ELL students comprised 20%.

Students with Specific Learning Disabilities. The SLD student data were retrieved from the total 28, 259 data files acquired from Districts A, B and C. The total number of student data files was 1,647 representing third grade students from 131 schools. Of this number, 1,542 student SLD files and 4,449 data points were utilized. Additional details about the eligibility and characteristics of the SLD data set are provided in this chapter. Table 4 summarizes the SLD data by district, race, and ethnicity. These data are compared to percentages of third grade student with learning disabilities in the three districts in 2006.
Table 4

Study Sample: SLD Membership by Race/Ethnicity Category

<table>
<thead>
<tr>
<th>Group Membership</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>District A</td>
<td>19% (37%)</td>
<td>54% (33%)</td>
<td>27% (27%)</td>
</tr>
<tr>
<td>District B</td>
<td>43% (57%)</td>
<td>29% (21%)</td>
<td>28% (19%)</td>
</tr>
<tr>
<td>District C</td>
<td>74% (80%)</td>
<td>6% (5%)</td>
<td>14% (12%)</td>
</tr>
</tbody>
</table>

*Note: Percentages in parentheses represent the racial/ethnic membership of the total student population in each district in 2006 as reported by the Education Information and Accountability Services, Florida Department of Education.*

Research Question 1 Variables

**English Language Learners.** The primary predictor variables of interest in the first research question are the ELL categories. Each ELL category provided a qualitative descriptor about ELL students therefore they were considered categorical variables.

Florida Statute §1003.56 (1990) defined ELLs as individuals who were not born in the US and whose native language was a language other than English; or as individuals coming from home environments where English is not the primary language. Although American Indians or Alaska natives are US born, they may be classified as ELLs if they are from an environment where a language other than English has had a significant impact on their English language proficiency. However, they were not included in the present study because they represented less than 3% of Florida’s ELL population.

Also, Florida Statute §1003.56 (1990) required ELLs to qualify or gain entrance into ELL or ESOL programs by demonstrating that they had sufficient difficulty speaking, reading, writing, or listening to the English language. Eligibility determination for ELL services began when students (and parents/caregivers) were surveyed upon entering a Florida public school and asked: (1) is a language other than English used in the home? (2) did the student have a first language other than English? and (3) does the
student most frequently speak a language other than English. An affirmative response to any one of these three questions automatically placed a student in an ESOL program where their status was listed as pending until a proficiency evaluation was conducted. Students in kindergarten through twelfth grade were assessed using a Florida Department of Education approved aural and oral language proficiency evaluation. If students’ scores fell within the limited English proficiency range, then these students were classified as ELL and were eligible for ESOL programs and or services. In contrast, students whose scores fell in the proficient range were ineligible for programs and or services. Is it noted that Florida districts follow English proficiency cut score recommendations that are outlined in standardized English proficiency test manuals.

An additional evaluation method occurred in third through twelfth grade where test scores from a normed-referenced test in reading and writing were examined. More specifically, ELL students were eligible for services if their standard scores on a state assessment fell at or below the 32nd percentile on a norm-referenced test. Finally, students were also eligible for ELL services if a school’s ELL committee (at the request of a parent or teacher) convened and examined data (e.g., academic and social observations or prior assessment records) that indicated a student should be considered for services. If an evaluation was recommended, students were eligible only if their proficiency results fell in the limited English proficient range (Florida Department of Education, 2005). In Florida, ineligible ELL students were categorized as Non-ELLs (or ZZ for Florida Department of Education data coding purposes). Non-ELLs represented students who (a) responded in the negative to three survey questions about English proficiency when seeking admission to a Florida public school or (b) responded in the
affirmative to one or more of the survey questions, but after an ELL assessment were not eligible for ESOL services.

The exit criterion for ELLs in kindergarten through second grade occurred when students obtained a Fluent English Speaker (FES) status on an approved proficiency assessment. For students in third grade or above, students were required to obtain a proficient level score of three or better on the Reading and or Writing portion of the Florida Comprehensive Assessment Test and be proficient in listening, speaking and writing based on a pre-approved English proficiency measure. Notwithstanding the aforementioned criteria, it is noted that an ELL committee may recommend that students continue in an ELL program if the majority of the committee believes and demonstrates that there is an academic need for continuation. Similarly, the ELL committee may recommend that an exited ELL student be re-admitted and re-classified into a program if evidence (e.g., grades and academic monitoring) suggest re-entry and ELL services are needed.

**Time.** The time variable in research question 1 represented benchmark data collected in the third grade during fall, winter and spring of the 2005/2006, 2006/2007 and 2007/2008 school years, respectively. Each assessment window was established by the Florida Department of Education’s Office of Assessment and School Performance who, in turn, communicated these timelines to each Florida school district. Next, district coordinators were responsible for communicating timelines to reading coaches at the school level; the reading coaches then coordinated data collection for third grades within the timelines established by the Office of Assessment and School Performance. Time was considered a categorical variable.
Gender. Gender was a categorical variable and was defined as male or female third grade students. Recent NAEP data indicated that gender differences in reading proficiency existed favoring girls at the elementary and middle school levels (National Center for Education Statistics, 2011). This trend continues in students’ post-secondary experiences even when factors like grade point averages, SES and race are controlled (Education Alliance, 2007).

Socio-economic status (SES). SES also was included in the ELL model. The present investigation categorized SES in two ways: (1) as students who were eligible and (2) as those who were ineligible for free and or reduced lunch. Although SES might be defined in many ways (e.g., occupation, level of education or income), data from a student’s free and/or reduced lunch status (FRL) is one way SES was determined in Florida. The Department of Agriculture’s National School Lunch Program determined a student’s free and/or reduced lunch eligibility by evaluating household income and occupants. Specifically, the National School Lunch Program determined that children qualified for free lunch if a household income of four people was not higher than 130% of the national poverty level. For a household of four, the reduced lunch eligibility income percentages ranged from 131% to 185% above the national poverty level. Florida districts comply with the Department of Agriculture’s National Lunch Program eligibility standards when qualifying students for their lunch programs.

Research Question 2 Characteristics and Variables

SLD Characteristic. The population of interest in research question 2 was students with learning disabilities. As discussed previously, the parameters for identifying a student with a learning disability were defined in the IDEA 1997 legislation
for state education agencies and their respective districts. Based on the guidelines specified in the law, districts and local education agencies (e.g., Child Study Teams) identified and qualified students with learning disabilities (a) using standardized assessment measures that revealed a significant discrepancy between one or more cognitive processes and an academic achievement measure; (b) when the team determined that underachievement was not due primarily to visual, hearing, motor, intellectual disabilities, emotional disturbance, or of environmental, cultural, or economic disadvantage; and (c) when specially designed instruction was needed to minimize the effects of underachievement (IDEA, 1997; National Research Center on Learning Disabilities, 2007). Table 5 highlights the federal law and demonstrates how Florida’s statute and the three districts’ (i.e., the districts associated with this study) procedural definitions were relatively consistent in their language regarding SLD determination.

Therefore, SLD students from the districts in the present study were students who were identified as having learning disabilities through processes that were consistent with IDEA, Florida Statutes and district requirements. The SLD data utilized in this study included students eligible for specialized instruction services in reading, math and writing because the three districts did not code data by specific learning disability content area. Also, the three districts coded all ineligible students for disability services under one generic coding category; therefore, a SLD comparison or control group was not possible. As a result, the investigator of this study compared the intercept and slopes of students with specific learning disabilities (e.g., characteristics that may manifest in an imperfect ability to listen, speak, read, write, spell, or do mathematical calculations) using time, race/ethnicity as coefficient variables.
### Table 5

**Definition of Specific Learning Disabilities per Federal Law, Florida Statutes and the three Districts’ procedures**

<table>
<thead>
<tr>
<th>National</th>
<th>State of Florida</th>
<th>District A</th>
<th>District B</th>
<th>District C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific learning disability is a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in an imperfect ability to listen, speak, read, write, spell, or do mathematical calculations. Such terms includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia and developmental aphasia. It does not include a learning problem that is primarily the result of visual, hearing, motor disabilities, mental retardation, emotional disturbance, or environmental, cultural, or economic disadvantage.</td>
<td>A specific learning disability is defined as a disorder in one or more of the basic learning processes involved in understanding or in using language, spoken or written, that may manifest in significant difficulties affecting the ability to listen, speak, read, write, spell, or do mathematics. Associated conditions may include, but are not limited to, dyslexia, dyscalculia, dysgraphia, or developmental aphasia. A specific learning disability does not include learning problems that are primarily the result of a visual, hearing, motor, intellectual, or emotional/behavioral disability, limited English proficiency, or environmental, cultural, or economic factors.</td>
<td>Specific learning disabilities refer to a heterogeneous group of psychological processing disorders manifested by significant difficulties in the acquisition and use of language, reading, writing, or mathematics. These disorders are intrinsic to the individual and may occur across the life span. Although specific learning disabilities may occur concomitantly with other handicapping conditions or with extrinsic influences, the disabilities are not primarily the result of those conditions or influences.</td>
<td>A specific learning disability is defined as a disorder in one or more of the basic learning processes involved in understanding or in using language, spoken or written, that may manifest in significant difficulties affecting the ability to listen, speak, read, write, spell, or do mathematics. Associated conditions may include, but are not limited to, dyslexia, dyscalculia, dysgraphia, or developmental aphasia. A specific learning disability does not include learning problems that are primarily the result of a visual, hearing, motor, intellectual, or emotional/behavioral disability, limited English proficiency, or environmental, cultural, or economic factors.</td>
<td>A specific learning disability is defined as a disorder in one or more of the basic learning processes involved in understanding or in using language, spoken or written, that may manifest in significant difficulties affecting the ability to listen, speak, read, write, spell, or do mathematics. Associated conditions may include, but are not limited to, dyslexia, dyscalculia, dysgraphia, or developmental aphasia. A specific learning disability does not include learning problems that are primarily the result of a visual, hearing, motor, intellectual, or emotional/behavioral disability, limited English proficiency, or environmental, cultural, or economic factors.</td>
</tr>
</tbody>
</table>
**Time.** In research question 2, the time variable represented assessment data collected during the same time periods discussed in research question 1. The assessment window procedures that were established by the Office of Assessment and School Performance also were identical. Like time in the ELL analysis, it was also considered a categorical variable in the SLD analysis.

**Race/Ethnicity.** The student race/ethnicity variables examined in research question 2 were White (Non-Hispanic), Black (Non-Hispanic), and Hispanic. Students identified as Asian, American Indian or Alaska Native, and Native Hawaiian or Pacific Islander were not included because their respective groups made up less than 3% of Florida’s student and state population. In 1997, the federal government’s Office of Management and Budget revised standards for the collection and reporting of race/ethnicity data; these standards and guidelines were implemented by the states and their respective districts. For the years 2006, 2007, and 2008, the Florida Department of Education reported and defined race/ethnicity based on six categories: (1) White: a person having origins in any of the original peoples of Europe, the Middle East, or North Africa; (2) Black or African American: a person having origins in any of the black racial groups of Africa; (3) Hispanic or Latino: a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race; (4) Asian: a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam; (5) American Indian or Alaska Native: a person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or
community attachment; (6) Native Hawaiian or Other Pacific Islander: a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands (Office of Management and Budget, 2000).

**Outcome Measure**

**The Dynamic Indicators of Basic Early Literacy Skills (DIBELS).** The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were developed at the University of Oregon to assess students’ development of early literacy skills in the areas of phonemic awareness, phonics, fluency, vocabulary and comprehension. Seven specific measures (i.e., Initial Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, Word Use Fluency and Retell Fluency) were developed to assess the areas. The purpose of these measures was to identify gaps in students’ early reading skills in order to provide teachers with direction on how to certify and intervene in problem areas. In effect, DIBELS were designed to be used in a prevention oriented decision-making framework such that potential reading problems were identified early and intervention was provided to prevent reading failure. The DIBELS oral reading fluency (DORF) was the primary measure that was used in this study.

The DORF was standardized and was administered individually to students in first through third grade. Students were required to read three standardized passages aloud for 1-minute each. At the end of one-minute, students were required to discontinue reading and their last word read was marked by a bracket. Examiners were instructed to subtract the total number of errors made by the student from the total number of words read during the one-minute assessment. This value was the number of words read correct
Correct words are those that were (a) pronounced correctly, (b) self-corrected within 3 seconds, and (c) words whose varying pronunciation might be explained by the local dialect or second language interference. Mispronounced words, substitutions, omitted words, words not pronounced correctly within the 3-second timeline, and word reversals were considered errors. An inserted or repeated word during the reading was not counted as a correct word or an error; it was ignored. The median score was used as the number of words read correct in one-minute and was recorded as the DORF measure (Good & Kaminski, 2002a).

Twenty-nine 3rd grade DORF probes of approximately equal difficulty were developed to facilitate the screening and progress monitoring of students. These probes were then evaluated for their readability (e.g., passage difficulty, accuracy of grade level, and length) using several readability formulas like Frye and Spache. The analyses compared readability formulas using nine readability indices that were pre-calibrated or weighted in the Micro Power & Light readability software. Statistical results revealed that the Spache readability formula was better suited to control for consistency of grade level passage difficulty. Specifically, stepwise regression analysis indicated that the Spache formula accounted for the greatest percentage of the variance (i.e., 30% or $r = .55$) in passage difficulty. In other words, using the Spache formula increased the chances that a reading probe was really at or near the grade level it was designed to assess (Good & Kaminski, 2002b).

The developers of the DIBELS present technical adequacy statistics on each of the five measures; however, only the technical adequacy of the third grade DORF is presented here. The DORF’s overall test-retest reliability (i.e., $r = .92$ to $r = .97$) and
alternate form reliability (i.e., $r = .89$ to $r = .94$) were reported as strong. Criterion related validity data were considered moderate to strong (i.e., $r = .52$ to $r = .91$). Additional DORF concurrent and predictive validity studies support the developers’ findings (Baker et al., 2008; Buck & Torgesen, 2002; Dynamic Measurement Group, 2008; Goffreda & DiPerna, 2010; Shaw & Shaw, 2002). A comprehensive review of these validation studies were discussed in chapter two.

The DORF reading probes were similar in nature to the oral reading probes used in CBM. For example, the DORF probes were about 200 words in length with an examiner’s copy and a corresponding student’s copy. The examiner’s copy had the cumulative word count at the end of every line on the right side of the page to facilitate scoring accuracy.

**Procedures**

The next section of this chapter described the DIBELS training and data collection procedures that were conducted in schools across the State of Florida. Next, the data retrieval processes at the district level and in relation to this study were outlined. Finally, a discussion about how the data were prepared and checked prior to conducting the present analysis concludes the procedures section.

**Procedures for Training and Administration.** Data collection training in the DIBELS was conducted by staff members from the Florida Center for Reading Research (FCRR). Training occurred in several phases for each Florida district. First, district level assessment teams were trained over a 2-day period in administration, scoring and reporting procedures. District team members included Reading First coaches assigned to Reading First schools, district level experts (e.g., reading specialists and school
psychologists) and key personnel (e.g., reading specialists and school psychologists) from Non-Reading First schools who were from schools identified as Title I or at-risk for reading failure. A second phase of training for district assessment team members required participants to practice administering and scoring the DIBELS screening materials with at least 20 students in grades K through 3 so that assessment proficiency was achieved. Advanced training was provided to district level team members before they facilitated training with school level teams. Specifically, district teams were trained on how to review data to make instructional decisions. For example, facilitators were trained how to determine if additional time and or supplemental instruction were needed to increase students’ outcomes in reading based on the score interpretation outlined in the Florida DIBELS School Readiness Uniform Screening System manual (FCRR, 2002; FCRR, 2005; Good & Kame’enui, 2002a).

The next phase of training occurred at the school level. School level personnel were provided with the same 2-day training and manual given to the district level facilitators; reading coaches typically facilitated the training at the school level. School level teams were typically comprised of support personnel (e.g., nurses, school psychologists, assistant principals, and media specialists). Familiarity and practice with scoring the paper and pencil version of the DIBELS was also a requirement for the local team.

Training using hand held computers was another level of training that was conducted for school-based teams. However, this technology and training were not part of the initial implementation in Florida schools. Specifically, the Dynamic Measurement Group and Wireless Generation (co-developers of DIBELS Palm), provided training
support on how to use the DIBELS Palm to district technology specialists and reading coaches. In turn, technology specialists, in collaboration with reading coaches provided training support to school-based teams on how to use the DIBELS palms. It was not known if every school or assessment team members within Reading and Non-Reading First schools used or were given standardized training in the use of the DIBELS palms.

Since the DIBELS had been administered to students, questions about its use with diverse populations (e.g., like ELLs and students with learning disabilities in the present study) were raised. The authors of the DIBELS reported that administering the DIBELS to ELLs and SLD students, among other groups, was appropriate because the goal for these students was to learn how to read in English and to monitor progress toward increasing benchmark goals (Dynamic Measurement Group, 2007). DIBELS administrators in Florida also addressed concerns related to its use with diverse learners. Specifically, questions about administering the DIBELS and its directions in a student’s native language were raised and addressed. Technical assistance for Florida educators outlined that accommodations were permitted for ELL students such that the administration directions could be given in a student’s native language (e.g., Spanish or Haitian Creole); however, an English time sensitive response was required. Translating and administering the English version of the DIBELS into another language was prohibited. When primary language directions were unavailable or an English response could not be elicited, students were excluded from testing (FCRR, 2005). As a reassurance to the integrity of this study, district research and evaluation personnel (i.e., from Districts A and C) reported that no ELL students were given a Spanish version of the DIBELS between the 2005/2006 through 2007/2008 school years. Therefore, the data
elicited from ELL students were either (a) entirely based on English directions and responses or (b) based on directions given in a students’ native language with an English response.

**Data Collection and School-wide Screening.** The reading coaches and other support staff at local schools collected the screening and progress monitoring data. Once data were collected during the prescribed assessment periods, *Reading First* districts (i.e., *Reading First* coaches) were required to upload their data for input to the Progress Monitoring Reporting Network (PMRN). PMRN is the web-based data collection and management system at the FCRR. In addition to data input and warehouse data storing, PMRN organizes student data so that the output is summarized and categorized to facilitate meaningful analyses, planning, and communication among students, teachers, and administrators. Non-*Reading First* districts and schools had the option of uploading data with the Progress Monitoring Reporting Network or storing their DIBELS data locally.

**Procedures for Data Retrieval.** A request to conduct the present study was submitted and approved by the Institutional Review Board of the University of South Florida (see Appendix A). Next, DIBELS data specific to third graders and their corresponding student characteristics (e.g., SES, gender, and education status) were requested for this investigation by completing district generated data request forms. A sample of one of these forms is located in Appendix B. Of the 67 requests made, three districts approved the request and provided the investigator with the relevant student data. Thus, the data for this study were first retrieved from the PMRN management system and then transferred to district computer servers by district personnel. Next, the requested
files were then retrieved by this investigator in three different ways. Specifically, District A mailed a compact disc containing their data via the US postal service, while data from District B were sent electronically and were password protected. A compact disc containing data from District C was collected by this investigator in person from the district’s research and evaluation office. Before the three districts sent any data, all identifiable information (i.e., names or social security numbers) were removed from the database. This investigator transferred and stored all data on a password protected computer.

**Procedures for Data Analysis Preparation.** Once the data were retrieved from the three districts, the next steps were to prepare the respective ELL and SLD data for analyses using Statistical Analysis Systems Version 9.3 (SAS 9.3). The following steps outline procedures for the ELL data. First, each district’s file was converted to a Microsoft Office Excel (2007) file if the files were delivered in a Statistical Product and Service Solutions (SPSS) format. Next, columns identifying a student as ELL or LEP were highlighted so that the column and all corresponding row information were extracted into a new Microsoft Excel spreadsheet. Districts A and C’s ELL data were then combined and ordered horizontally by year (i.e., 2006, 2007, and 2008), district letter, school and student number, and then by predictor (i.e., ELLs, SES and Gender categories) and outcome variables (i.e., benchmark assessments). Third, the predictor variables were changed from their original numeric or alphabetic designations and dummy coded. Specifically, the ELL groups were dummy coded in the regression analysis as follows: LY = 1 or LY = 0; LF = 1 or LF = 0; and LZ = 1 or LZ = 0 such that a 1 indicated the student was a member of the designated group and a 0 indicated the
student was not a member of that group, and as a consequence the presence of a 0 for all three dummy coded variables indicated a Non-ELL student and a reference group member. Fourth, the SES and gender predictors were dummy coded similarly. Relative to SES, the free and/or reduced lunch data was treated as a categorical variable and coded as SES-Lunch = 1 or SES-Lunch = 0. SES-Lunch = 1 represented third grade students who qualified for free and/or reduced lunch and SES-Lunch = 0 represented those third grade students who were ineligible for free and/or reduced lunch. Gender regression coefficients were entered into the model as gender = 1 or gender = 0, where females served as the reference group and were coded as 0 while males were coded as gender = 1.

Finally, Time also was dummy coded as a categorical variable and was coded with two dummy variables (i.e., Winter and Spring) such that the Fall administration represented the reference time. More specifically, Winter = 0 and Spring = 0 implied fall data, while Winter = 1 and Spring = 0 implied winter data, and Winter = 0 and Spring = 1 implied spring data. Thus, the question 1 predictor and outcome variables that were identified for the SAS 9.3 analysis were the ELL (i.e., LY, LF and LZ) and Non-ELL categories, gender, SES and Time (representing Fall, Winter and Spring DIBELS scores across 2006, 2007, and 2008).

The students with disabilities data were identified utilizing the exact process used with the ELL data set; however, instead of highlighting the LEP data column, the ESE (i.e., exceptional student education) column and their respective information rows were highlighted in the original dataset. Once all ESE columns and row data were extracted and placed into a new Microsoft Excel spreadsheet, the data were again highlighted to extract only those columns and rows with a letter K designation. In Florida schools, the
letter K was the data entry code representing a student with a learning disability. A final tiered extraction required the investigator to identify the race/ethnicity column, then highlight the alphabetic designations of A, M, and I. These letters represented Asian, Multi-or Bi-racial and Indian students, in that order. These files were eliminated from the SLD file leaving only W, B, and H letter codes which represented White, Black and Hispanic race/ethnic categories, respectively.

Also, like the ELL data file, all three SLD district files were combined and ordered horizontally to reflect district year, letter, school, student number, and then predictor and outcome variables. The race/ethnicity predictors were assigned two betas to reflect the presence and absence of the type of race or ethnicity. The two dummy coded variables were Black (0 = Non-Black, 1 = Black) and Hispanic (0 = Non-Hispanic, 1 = Hispanic), and thus students that were White and Non-Hispanic served as the reference group (i.e., Black = 0, Hispanic = 0). Time was coded with two dummy variables (Winter and Spring) such that the fall administration represented the reference time. More specifically, Winter = 0 and Spring= 0 implied fall data, while Winter = 1 and Spring = 0 implied winter data, and Winter = 0 and Spring = 1 implied spring data. Once the predictor variables (i.e., White, Black, and Hispanic and Time) were dummy coded, the SLD data files were ready for analysis using the SAS 9.3 software.

Data Reliability Checks. In a quasi-experimental design, inter-rater reliability checks of the data were analogous to inter-rater reliability in a randomized experiment design. The purpose of the data reliability checks was to assure the presence and consistency of the data across two independent checkers. The inter-rater reliability check process involved several steps. Specifically, the process began with a random sampling
of the ELL and SLD student files. Next, the training steps related to the ELL and SLD data were outlined.

First, the ELL and SLD student files were subjected to a simple random sampling procedure. Simple random sampling (SRS) was a selection process where every ELL and SLD data file had an equal chance of being selected (Urbaniak & Plous, 2011; Yates, Moore, & Starnes, 2008). For the ELL data files, the Research Randomizer (a computer-based random number generator) was employed to generate 2700 random numbers between 1 and 27,000 which represented 10% of the ELL data. Next, the Research Randomizer generated 1647 random numbers representing the SLD data files; of this total, the first two hundred files or 12% of the SLD data files were prepared for inter-rater reliability checks. It was noted that each number generated for the ELL and SLD data files became the Microsoft Excel spreadsheet number (located to the left of the first column) that was used to identify a student file by the data evaluators.

The next step related to the data reliability checks were the two brief training (i.e., a range of 40 to 60 minutes) sessions. These sessions were held by the investigator on two separate occasions; the first session with two school psychology graduate students and the second session with two district level educators (one a school psychologist and the other a data analyst/trainer in District C). The purpose of the training was to demonstrate how the reliability check protocols were to be scored (see Appendices D and E). Therefore, the investigator demonstrated the steps needed to correctly conduct a data file check during the training. First, data checkers were required to open the prepared data file and the original data file in Microsoft Excel. The data evaluators were shown how to use the “view side by side” option located within the View folder of the Microsoft
Excel program. This option allowed checkers to view the prepared and original data file simultaneously. Next, checkers opened another Microsoft Excel spreadsheet folder in order to retrieve the next randomly generated number from the ELL or SLD randomized number list. This number represented the prepared data list’s Microsoft Excel line item number located just before column one on the spread sheet. Once this number was located on the prepared data spread sheet, the data checkers then were instructed to locate the student identification that corresponded with the line item number; the student number was located under the student identification column. Next, the student identification number was used to obtain the corresponding district letter code (i.e., District A, B or C), and the assessment year (located by column) within the prepared data file. The student number, the district identifier and the assessment year from the prepared data file were then matched with corresponding data from the original data file. To facilitate this step, checkers were instructed to insert the student number in the “find” option drop box in the Microsoft Excel computer program so that the computer sorted and identified the student files that needed to be cross checked. Data evaluators then were shown how to check for the presence of school identification, and school grade information. Next, the second of the three columns containing the winter DORF score was located. The presence or absence of this score was then matched with the second column containing the DORF score in the original data file.

Specific to the ELL data files, data checkers then located the three ELL categories columns (i.e., ELL-LY, ELL-LF, and ELL-LZ) on the ELL prepared Microsoft Excel spread sheet and then the corresponding LEP (Limited English Proficient) column on the ELL original Excel spread sheet in order to examine if the dummy codes on the prepared
data spread sheet corresponded with the district ELL coding (i.e., LY, LF, LZ or ZZ) structure. For instance, a zero in all three columns on the ELL prepared spread sheet meant that the corresponding LEP column in the original data file should have contained a /ZZ/, indicating that the student file being examined represented a Non-ELL student. A check was placed on the ELL data check form if the data checker found that there was consistency in the ELL coding. Evaluators were provided with the dummy codes and their corresponding ELL district codes.

Next, reliability checkers crossed checked the gender coding accuracy. On the prepared Excel spread sheet the gender column included a zero or a one; a zero represented females, a one represented males. On the original data file, the gender columns were either labeled “gender” or “sex” to signify the gender of the students. Cross checkers provided a check if a zero on the prepared list was represented by a /F/ signifying a female student in the original ELL data file. Similarly, checkers placed a check on the ELL data form if a one was located in the gender column of the prepared data file and a corresponding /M/ or the word Male was found in the gender column of the original ELL data files. This process was repeated for verification of the free and/or reduced lunch data. Specifically, the columns on the prepared ELL data file and original ELL data files that represented free and/or reduced lunch data (SES in the prepared data Excel files and Lunch or RFrl in the original data file) were checked by the checker as accurate if a one on the prepared data files corresponded to a /Y/ or the word yes in the original data Excel data files. Also, checks were given if a /N/ or the word no was found in the Free and/or Reduced Lunch column of the original ELL data file and corresponded
to a zero in the SES column of the prepared data file. A copy of the ELL data check form is located in Appendix D.

The SLD checkers verified the presence of the SLD category on the prepared data file by ensuring that a /K/ was in the exceptional student education or ESE column. Next, data check evaluators cross checked the prepared and original data files to determine the presence of the second ORF score; checkers verified the ORF presence with a check mark. Next, the student’s race/ethnicity status was cross checked. In the original data set, District A and C used the label ‘race’ as their column identifier. The student’s race was either spelled out as White, Black, Hispanic or was identified with the letters /W/, /B/, /H/ representing the aforementioned race/ethnicity, in that order. Checkers verified this information by examining the dummy coding across two race/ethnicity columns (Black and Hispanic) in the prepared data file. For instance, if the sequence across the Black and Hispanic columns was a one then a zero, the race/ethnic column in the original data file should have been represented by the word Black or the letter/ B/. Similarly, if the sequence across the Black and Hispanic columns was a zero then a one, the race/ethnic column in the original data file should have indicated the word Hispanic or the letter /H/. A sequence in the prepared data race/ethnicity column that was zero followed by another zero meant that the student race/ethnicity category was listed as White or the letter /W/ in the original data files. Following the race ethnicity checks, data checkers then verified that the student identification number across both files corresponded with the SLD district code signified by the letter /K/.

Once the training session ended, confidentiality and protection forms were signed (see Appendix C). The investigator was available for additional consultation by phone
after the training. One rater received additional consultation by phone. Graduate assistants and the district educators returned the data check forms to the study investigator who, in turn, conducted the inter-rater reliability calculations. Inter-rater reliability was calculated using the following formula:

\[
\frac{\text{Total Number of Correct Data Checks}}{\text{Total Number of Correct Data Checks + Total Number of Incorrect Data Checks}} \times 100
\]

An overall 90% accuracy rate was the minimum criteria needed for the data accuracy procedure. Ninety percent accuracy suggests a high degree of assurance that the consistency between reliability checkers was strong (Einfeld, Tonge, Chapman, Mohr, Taffe, & Horstead, 2007). The overall inter-coding reliability for the ELL students’ files was 100%. This total represented .16% of the 26,967 data files. Inter-coder reliability for the SLD data files was 96%, representing 3% of the 1542 student data files.

**Study Design**

The present investigation employed a longitudinal, quasi-experimental research design to answer the research questions. The study was considered longitudinal in nature because it used individual observations at three different points in time. Also, the data gathered comprised observations of students from across three consecutive years (i.e., 2006, 2007, and 2008). The study was considered quasi-experimental in design because group comparisons were made at some levels, even though the archival nature of the study prevented the random assignment of participants. The technical adequacy of the study was enhanced through randomization of the data check process. The data check process was specified in Appendices D and E.
Data Analysis Procedures

The next section of this chapter described the statistical analyses employed in the study. Specifically, the descriptive and inferential statistics analyses utilized were discussed.

Descriptive Analysis. In addition to summarizing the demographic data using frequency tables, the mean and standard deviations (i.e., measures of central tendency and variability) were reported for the dependent variables in the present investigation.

Hierarchical Linear Modeling. To answer the inferential portions of the research questions, a two-tier multi-level analysis was employed. Tiered multiple regression analysis is also known as hierarchical linear modeling (HLM). The term “hierarchical” is used because the data are subjected to an incremental step order that proceeds from the micro- to the macro-level (Luke, 2004; Yaffee, 1996). HLM facilitates the examination of multi-level systems (micro and macro) simultaneously holding variables within and between levels constant in order to determine the effect on the outcome variable(s). In effect, the goal of HLM is to predict the value of the dependent variable based on one or more predictor variables (Luke, 2004; Shay & Gomez, 2002; Woltman, Feldstain, MacKay, & Rocchi, 2012). Growth curve analyses (GCA) is a modification of the HLM procedure that examines growth over time. Thus, one of the predictor variables used in this investigation was time.

Multi-level systems analyses are often also needed because data are typically nested or clumped together, violating the independent error assumption associated with correlation design (Raudenbush & Burk, 2002; Yaffee, 1996). Specific to the present study, micro-level data consisted of individual observations and/or student data that were
nested within individuals at the macro-level of the model. In essence, the HLM model allowed each level in the structure to be formally represented by its own sub-model without compromising statistical power significantly. In turn, the sub-model expressed relationships among variables within a given level, and specified how variables at one level interacted or influenced variables at other levels (Raudenbush & Bryk, 2002; Woltman, Feldstain, MacKay, & Rocchi, 2012).

One advantage of using the HLM procedure with the present investigation was that the HLM model effectively accommodated missing data (Luke, 2004; Shay & Gomez, 2002). A disadvantage of using a multi-level model approach, as with other statistical approaches, was that the HLM procedure required parameters that increased or maximize the fit of the model, while decreasing the amount of residual. To do this, it was necessary to include variables that were not of primary interest to this investigation. For example, the present investigation also examined the relationship of socio-economic status (SES) on DORF outcomes.

In the current study, the two-tier multi-level growth curve analyses were analyzed using the Statistical Analysis System (SAS) Version 9.3. Research question 1 asked:

To what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (ELLs) (a) enrolled in an ESOL program (LY), (b) not enrolled but monitored (LF), (c) not enrolled and no longer monitored, and non-ELLs when holding gender and SES constant?

Research Question 2 inquiry was:

To what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnic groups?
For each question, the investigator first employed an unconditional or unconstrained model where the intercept or initial average status was identified. The unconditional model was represented as $DIBELS_{ij} = \pi_{0j} + e_{ij}$. The unconditional model was calculated because it determined if there were any significant differences between group estimates that needed further exploration. In effect, the unconditional model allowed the researcher to calculate the interclass correlations (ICC). The ICC measured the proportion of the variance that was accounted for within and between groups.

Next, the model construction continued with the Level 1 equation: $DIBELS_{ij} = \pi_{0j} + \pi_{1j} * Time_{ij} + e_{ij}$ where $\pi_{0j}$ represented the intercept or expected outcome for observations resident in each student; $\pi_{1j}$ represented the predicted slope or growth based on time as a predictor. Next, Level 2 of the model represented student characteristics (i.e., gender and SES) in research question one. A regression equation for Level 2 was represented as follows:

$$\pi_{0j} = \beta_{00} + \beta_{01} * Gender_j + \beta_{02} * SES_j + \beta_{03} * ELL\_LY_j + \beta_{04} * ELL\_LF_j + \beta_{05} * ELL\_LZ_j + \mu_{0j}$$

$$\pi_{1j} = \beta_{10} + \beta_{11} * Gender_j + \beta_{12} * SES_j + \beta_{13} * ELL\_LY_j + \beta_{14} * ELL\_LF_j + \beta_{15} * ELL\_LZ_j + \mu_{1j}$$

The intercept and slopes were grand mean or zero centered. It was hypothesized or expected that random effects on the intercept and the slopes would be observed at Level 2 because students would likely have different reading skills at the beginning and would also vary in their reading growth over time. This portion of the GCA procedure allowed the investigator to examine the variability within and between individual characteristics using the grand mean as a comparison (Yaffee, 1996; Levels of Analysis, 2006). Finally,
fixed and random estimates were also generated and discussed for all the variables (i.e.,
ELLs, Gender and SES) in the model.

When Levels 1 and 2 equations were combined, the model became:

\[
DIBELS_{ij} = \beta_{00} + \beta_{01} \cdot \text{Gender}_j + \beta_{02} \cdot \text{SES}_j + \beta_{03} \cdot \text{ELL} \_ \text{LY}_j + \beta_{04} \cdot \text{ELL} \_ \text{LF}_j + \\
\beta_{05} \cdot \text{ELL} \_ \text{LZ}_j + \beta_{10} \cdot \text{Time}_{ij} + \beta_{11} \cdot \text{Gender}_j \cdot \text{Time}_{ij} + \beta_{12} \cdot \text{SES}_j \cdot \text{Time}_{ij} + \\
\beta_{13} \cdot \text{ELL} \_ \text{LY}_j \cdot \text{Time}_{ij} + \beta_{14} \cdot \text{ELL} \_ \text{LF}_j \cdot \text{Time}_{ij} + \beta_{15} \cdot \text{ELL} \_ \text{LZ}_j \cdot \text{Time}_{ij} + e_{ij} + \mu_{0j} + \\
\mu_{ij} \cdot \text{Time}_{ij}
\]

Similarly, the model for research question two added time as a predictor after an
unconditional model and ICC scores were constructed and derived: \(DIBELS_{ij} = \pi_{0ij} + \pi_{1ij} \cdot \text{Time}_{ij} + e_{ij}\). Essentially, the intercepts and slopes examined if students with learning
disabilities demonstrated differences based on initial reading skills alone, or were there
noted differences based on growth over time as well. Next, Level 2 of the model
represented observations nested in students using race/ethnicity as predictors. A
regression equation for Level 2 was represented as follows:

\[
\pi_{0ij} = \beta_{00} + \beta_{01} \cdot \text{Black}_j + \beta_{02} \cdot \text{Hispanic}_j + \mu_{0j}
\]

\[
\pi_{1ij} = \beta_{10} + \beta_{11} \cdot \text{Black}_j + \beta_{12} \cdot \text{Hispanic}_j + \mu_{1j}
\]

Like the ELL model, the SLD model intercept and slopes were grand mean centered.
Random effects on the intercept and the slopes at Level 2 were expected because students
varied in their initial reading skills at the beginning and also varied in their reading
growth over time.

When the levels associated with research question two were combined, the following
equation was derived:

\[
DIBELS_{ij} = \beta_{00} + \beta_{01} \cdot \text{Black}_j + \beta_{02} \cdot \text{Hispanic}_j + \beta_{10} \cdot \text{Time}_{ij} + \beta_{11} \cdot \text{Black}_j \cdot \text{Time}_{ij} + \\
\beta_{12} \cdot \text{Hispanic}_j \cdot \text{Time}_{ij} + e_{ij} + \mu_{0j} + \mu_{ij} \cdot \text{Time}_{ij}
\]
Chapter Four

Results

This chapter reports the findings of the two research questions associated with the present study. The study’s research questions asked (1) to what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (ELLs) enrolled in an ESOL program (LY), ELLs not enrolled but monitored (LF), ELLs not enrolled and no longer monitored (LZ), and non-ELLs when holding gender and SES constant? and (2) to what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnic groups? The chapter is organized into three sections. First, ELL descriptive and mixed modeling results, along with assumption analyses are reported. Second, a similar organizational style is used to report the results obtained from the analyses of the SLD data. Finally, the chapter concludes with a summary of the major findings.

Research Question 1

Descriptive Analysis. Before hierarchical linear models were estimated for the research question involving ELL students, descriptive analyses were conducted on the 26,967 third graders and their respective 281 schools from Districts A and C. As reported in chapter three, ELL students were represented in three sub-categories in this study: ELL-LY, ELL-LF, and ELL-LZ. ELL-LY represented students who were eligible and enrolled in programs and or services designed for ELL students. ELL-LF represented students who were monitored for two years after exiting an ELL program or service and
ELL-LZ represented students who exited ELL programs and who were no longer monitored. Non-ELLS (a comparison group) were students who were determined to be Native English speakers (based on a state-mandated survey that was given to all students and parents upon entrance in a Florida public school) and therefore were not eligible for ELL services. Descriptive findings revealed that the ELL students comprised 20% of the data sample, and the sample was equitable for males and females. Socio-economic status data revealed that 69% of the students in the sample qualified for free and/or reduced lunch. Table 6 provides a summary of the ELL and Non-ELL groups by district.

Table 6

<table>
<thead>
<tr>
<th>Study Sample: Demographic Data of ELL and Non-ELL Groups by District</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Membership</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>District A (82%)</td>
</tr>
<tr>
<td>District C (18%)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Note: Total N = 26,967.

Independent of group association, ELL and Non-ELL third grade students were individually administered the DORF measures for each year data were obtained (i.e., 2006, 2007 and 2008). During each assessment period, students read three standardized passages aloud for 1-minute each. The number of words read correct for each passage was calculated and the median of the three scores obtained was considered the student’s DORF measure for that assessment period (Good & Kaminski, 2002a). The total DORF measures or observations used in the analysis were 61,248, representing 20,025 student files.
A visual inspection of the means suggested that students’ performance increased overtime. Skewness and kurtosis values were near zero, suggesting that the sample of ELL and Non-ELL students was normally distributed. It was noteworthy that observations from the Spring DORF administrations were significantly less than the observations from the Fall and Winter administrations. Missing and or unrecorded data were the likely reasons for the numeric differences. Table 7 provides a descriptive summary of the ELL data set.

Table 7

<table>
<thead>
<tr>
<th>Time</th>
<th>n⁸</th>
<th>Mean (SD)⁹</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>26576</td>
<td>74.18 (34.01)</td>
<td>.19</td>
<td>.03</td>
</tr>
<tr>
<td>Winter</td>
<td>23981</td>
<td>98.26 (36.01)</td>
<td>.02</td>
<td>.30</td>
</tr>
<tr>
<td>Spring</td>
<td>11264</td>
<td>99.27 (34.62)</td>
<td>-.01</td>
<td>.49</td>
</tr>
</tbody>
</table>

Note: ⁸ = number of observations collapsed across 2006, 2007, 2008; ⁹ = standard deviation.

**ELL Multi-Level Model Results.** In contrast to the descriptive data, growth curve or mixed modeling procedures were utilized to design a two-level confirmatory approach model. In effect, prior established research about ELLs, the DIBELS and their respective correlates, along with some unanswered questions in the research, formed the basis for the research question and the structure of the ELL model. The following research question was asked: To what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (ELLs) enrolled in an ESOL program (LY), ELLs not enrolled but monitored (LF), ELLs not enrolled and no longer monitored (LZ), and non-ELLs when holding gender and SES constant? The growth curve model constructed to answer this question follows immediately:
\[
\text{DIBELS}_{ij} = \beta_{00} + \beta_{01}\text{Gender}_i + \beta_{02}\text{SES}_i + \beta_{03}\text{ELL\_LY}_j + \beta_{04}\text{ELL\_LF}_j + \\
\beta_{05}\text{ELL\_LZ}_j + \beta_{10}\text{Time}_ij + \beta_{11}\text{Gender}_i\text{Time}_ij + \beta_{12}\text{SES}_i\text{Time}_ij + \\
\beta_{13}\text{ELL\_LY}_j\text{Time}_ij + \beta_{14}\text{ELL\_LF}_j\text{Time}_ij + \beta_{15}\text{ELL\_LZ}_j\text{Time}_ij + e_{ij} + \mu_{ij} + \\
\mu_{0j} + \text{Time}_ij
\]

The two-level model converged using fixed and random predictors. Relative to
the fixed effects, the intercept findings revealed that the grand mean was 87.15. This
score represented the reference group’s (i.e., Non-ELL females who were not eligible for
free or reduced lunch) initial value on the DORF outcome when predictors were zero
centered. When the intercept was compared to the ELL subgroups, initial level results
varied. ELL-LY students’ initial levels (Estimate = -16.67, \(p < .05\)) were negative and
significantly lower than the reference group. In contrast, the ELL-LZ students’ results
(Estimate = 7.79, \(p < .05\)) were positive and significantly higher than students from the
reference group. The estimate (Estimate = .60, \(p > .05\)) observed in the ELL-LF group
was positive but not significant (see Table 8).

Time also was a significant predictor of students’ DORF outcomes suggesting
that the average DORF scores increased significantly over time. Thus, for each unit
increase in time (typically every three months in the school year) the DORF scores of
Non- ELL female students who were ineligible for free and or reduced lunch increased by
16.58 points from the initial average of 87.15. Similar to the positive initial slope
estimate, the three ELL group’s slopes were also positive. It is noted that the ELL-LY
group’s slope (Estimate = 2.46, \(p < .05\)) was significantly higher than the reference group
(see Table 8).

Gender and SES were also modeled as fixed effects and were considered constant
predictor variables. Initial level findings for these two groups were negative and
significantly lower than the reference group’s DORF score (Gender Estimate = -6.31,
SES = -8.26, $p < .05$). Significant and negative results were found for the gender slope (Gender Slope Estimate = -1.03, $p < .05$) as well. In contrast, the SES slope was positive and significant (SES Slope Estimate = 1.17, $p < .05$). When placed in the context of the present analyses, the findings suggested that when compared to the reference group, students on free and reduced lunch were at a disadvantage relative to their initial average DORF scores; however, their slope score was higher than the average slope, which might result in greater DORF gains over time. Although the present ELL model only provided significance tests comparing each ELL group to the reference group, predicted intercept and slope values were derived for ELL and Non-ELL students using sixteen group combinations. A graphical representation (see Figure 1) of these predicted lines provides a sense of the possible influence that predictors (e.g., SES) can have on diverse learners. For example, ELL-LY males who received free and or reduced lunch appeared to be at greatest risk for not meeting reading fluency benchmarks at the end of their third grade year.

The variance parameters estimates were also examined. Findings for the covariance estimate (i.e., between the intercept and the slope residuals) was significant (Estimate = 143.05, $p < .05$); however, the variance estimate for the time slope residuals was zero. A zero value for the slope residual variance might be indicative of over specification of the model, or simply that there was no random variation within the ELL sample population for the time effect. The value of the Level 1 residual variance estimate was 472.11, $p = .0001$. This estimate was smaller than a preliminary investigation of the unconditional model’s estimate of 698.70. In general, smaller residuals are not
indicative of over specification, but are indicative of appropriately fitted predictors in a model. Variance estimate totals are presented in detail in Table 9.

Table 8

\textit{Growth Curve Model: Prediction of DORF Estimates for English Language Learners}

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>T</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textbf{Initial Levels}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>87.15</td>
<td>.36</td>
<td>240.37</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL-LY</td>
<td>-16.67</td>
<td>.53</td>
<td>-31.44</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL-LF</td>
<td>.60</td>
<td>.68</td>
<td>.89</td>
<td>0.37</td>
</tr>
<tr>
<td>ELL-LZ</td>
<td>7.78</td>
<td>.67</td>
<td>11.05</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>-6.31</td>
<td>.31</td>
<td>-20.24</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>SES</td>
<td>-8.26</td>
<td>.35</td>
<td>-23.69</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>\textbf{Slopes}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>16.58</td>
<td>.26</td>
<td>63.85</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL-LY*Time</td>
<td>2.46</td>
<td>.44</td>
<td>5.59</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL-LF*Time</td>
<td>0.92</td>
<td>.55</td>
<td>1.64</td>
<td>.1012</td>
</tr>
<tr>
<td>ELL-LZ*Time</td>
<td>0.37</td>
<td>.55</td>
<td>.67</td>
<td>.5055</td>
</tr>
<tr>
<td>SES*Time</td>
<td>1.17</td>
<td>.27</td>
<td>4.39</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Gender*Time</td>
<td>-1.03</td>
<td>.25</td>
<td>-4.14</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: \textit{p value is significant at } p < .05.

Table 9

\textit{Growth Curve Model: Variance Estimates for English Language Learners}

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>532.91</td>
<td>9.26</td>
<td>57.50</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Covariance</td>
<td>143.05</td>
<td>4.85</td>
<td>29.45</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>472.11</td>
<td>3.32</td>
<td>142.22</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: \textit{p value is significant at } p < .05.
Figure 1. Predicted DORF trajectories for 16 Non-ELL and ELL groups. The graph illustrates the trend for male (M) and female (F) students with (W/) and without (W/O) free and/or reduced lunch (FRL).
**Assumptions.** Investigating the robustness of the ELL data to the assumptions associated with growth curve analysis was a pre-requisite to constructing the model. Specifically, a priori inspections relative to normality, linearity, homoscedasticity, and multicollinearity assumptions were examined. The respective analyses were conducted utilizing a random sub-sample of the ELL data because the 61,248 observations in the original ELL data set were too large for the SAS 9.3 software to generate the output requested by the Mixed_DX macro. Simple random sampling (SRS) was the random process used to extract a sub-sample of the ELL data, thus allowing each student’s identifier an equal chance of being selected. The Mixed_DX macro was a comprehensive syntax or coding that was used to examine assumptions associated with a two-level linear model structure simultaneously (Bell, Morgan, Schoeneberger, Ferron, & Kromrey, 2009; Yates, Moore, & Starnes, 2008). Thus, 2700 students’ identification numbers or 8100 observations were randomly sampled using the SAS 9.3 software survey select procedure; these numbers represented 10% of the ELL data file. However, because of missing data, the SAS 9.3 software only utilized 6117 observations. The final sub-sample total represented 7.6% of the ELL data set.

The assumption process began by examining normality. The normality assumption required the investigator to determine if the residuals were normally distributed. Significant violations to the normality assumption sometimes bias estimation effects leading to an inappropriate acceptance or rejection of the null hypothesis (Osborne, 2000; Woltman et al., 2012; Yu, 2011). However, when mixed modeling procedures have sufficient power (i.e., large group and sample sizes) they are typically robust to normality violations (Raudenbush & Byrk, 2002). In the present model,
normality assessments revealed approximately normal distributions. Specifically, inspections of skewness and kurtosis values were .17 and 3.08, respectively. Examination of the skewness and kurtosis values provided sufficient information about normality because values near zero suggested a normal distribution, and values significantly greater than zero indicated the presence of extreme outliers. The linearity assumption appeared to be supported as well. Linearity suggests a straight line or form relationship, either negative or positive, between variables under investigation. Yu (2011) recommends using residuals when evaluating linear relationships because residuals control or account for the partial relationships among all the variables under investigation. Visual inspection of the residual scatter plot revealed a linear relationship between the residuals and predicted values (see Figure 2).

![Plot of level-1 residuals/predicted values](image)

**Figure 2.** ELL scatter plot of residuals and predicted DORF scores.

Visual inspection also was the method employed to determine the extent to which the homoscedasticity assumption was justified. Homoscedasticity is concerned with the constancy of variances (Yu, 2011). In the ELL model, the scatter plot in Figure 2 above
demonstrated that heavy concentration of the residuals was centered near zero with no particular pattern. Also, the sparse numbers of residuals above and below the mean of zero were indicative of outliers, but they did not appear to disrupt the constancy of the variances that were at or near the zero point.

Prior to estimating the ELL model, a multicollinearity assessment was also conducted. Multicollinearity is concerned with high inter-correlation among variables. The presence of collinearity is indicative of model imprecision and is the result of two or more variables being highly correlated. To determine the presence of collinearity among the predictor variables (i.e., Time, ELL subgroups, gender and SES), the variance inflation factor (VIF) was calculated using the SAS 9.3 regression procedure. The VIF calculates the level of inflation in the dataset. Yu (2011) reports that when the VIF statistics are at or below 10, collinear absence is assumed. The range VIF in the present model was between 0 and 1.018.

Also, inspection of some of the 61,248 observations revealed that some student data files had as much as 9 observations (i.e., DIBELS scores) over the 3-year period. A maximum of 3 observations per student number was expected if a student completed one year in third grade. In effect, there were suspicions that some of the observations came from students who completed more than one year in third grade. Therefore, the data set was re-sorted to identify and eliminate observations that were more than 3 per student identifier, and then the ELL model equation was re-calculated in SAS. Findings indicated that the ELL groups’ initial (Estimate = 88.26, p < .05) and slope (Estimate = 17.91, p < .05) estimates were similar to those previously obtained.
Research Question 2

**Descriptive Analysis.** The second research question asked: to what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnic groups? As mentioned in chapter three, there were 1647 data files retrieved from the larger data set that categorized participants as students with specific learning disabilities. A summary of this sample is presented in Table 10.

Table 10

<table>
<thead>
<tr>
<th>District (Number of Schools)</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Race/Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
</tr>
<tr>
<td>District A (81)</td>
<td>850</td>
<td>161 (19%)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>District B (5)</td>
<td>119</td>
<td>51 (43%)</td>
</tr>
<tr>
<td>District C (43)</td>
<td>678</td>
<td>526 (78%)</td>
</tr>
<tr>
<td>Total (131)</td>
<td>1647</td>
<td>738 (45%)</td>
</tr>
</tbody>
</table>

*Note: <sup>a</sup> = Number of observations, <sup>b</sup> = Percentage equivalent in each group.*

Similar to the ELL outcome data, the DORF was individually administered to third grade students in each respective year 2006, 2007 and 2008. The assessment administration protocol was the same for the SLD data sample. Table 11 provides summary statistics of the Time variables associated with Fall, Winter, and Spring, respectively.

Table 11

<table>
<thead>
<tr>
<th>Time</th>
<th>N</th>
<th>Mean (SD)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>1568</td>
<td>35.14 (25.69)</td>
<td>.83</td>
<td>.65</td>
</tr>
<tr>
<td>Winter</td>
<td>1516</td>
<td>47.84 (31.31)</td>
<td>.45</td>
<td>-.19</td>
</tr>
<tr>
<td>Spring</td>
<td>1415</td>
<td>55.71 (34.75)</td>
<td>.34</td>
<td>-.37</td>
</tr>
</tbody>
</table>

*Note: N = number of observations. SD = standard deviation values.*
A review of the means across each time point indicated change over time, and the skewness and kurtosis values (i.e., greater than or less than zero) suggested an approximately normal SLD sample distribution. However, mixed model procedures have more precise estimates when measures of variability are calculated using residuals (Raudenbush & Byrk, 2002; Yu, 2011).

**SLD Multi-Level Model Results.** Like the ELL model, the SLD model was constructed using a confirmatory analysis approach. Decisions about fixed and random variables were determined from the research literature about students with disabilities, their race/ethnicity, and their respective performance on outcome measures like the DORF. Thus, the following mixed model equation was derived:

\[
DIBELS_{ij} = \beta_{00} + \beta_{01} \times \text{Black}_j + \beta_{02} \times \text{Hispanic}_j + \beta_{10} \times \text{Time}_{ij} + \beta_{11} \times \text{Black}_j \times \text{Time}_{ij} + \\
\beta_{12} \times \text{Hispanic}_j \times \text{Time}_{ij} + e_{ij} + \mu_{0j} + \mu_{1j} \times \text{Time}_{ij}
\]

Time, Black and Hispanic served as fixed coefficients, while time and the intercept (both zero centered) were allowed to covary. The convergence criteria were met providing covariance and fixed coefficients estimates. Statistically significant findings for the intercept showed that the initial DORF reading estimates were higher \((p < .05)\) for the reference group (i.e., White students with SLD). Slope or growth estimates revealed that time was a significant predictor of DORF measures (Estimate = 13.85, \(p < .05\) for the White group); however, there were no significant differences between the reference group’s slope (i.e., White students) and the slope over time of the Hispanic or Black groups (Hispanic slope difference from White = -.24, \(p > .05\); Black slope difference from White = -.69, \(p > .05\)). Comprehensive statistical findings for the SLD groups are found in Table 12. A linear graph, presented in Figure 3, demonstrates what these
estimates may look like for the present sample of students with learning disabilities when considering their three distinct race/ethnic backgrounds.

The variance parameters estimates were also examined. Findings for the covariance estimate (i.e., between the intercept and the slope) was significant (Estimate = 95.71, \( p < .05 \)); however, a zero value for the slope effect was found. This value suggested that there was no variation in the model for the time effect. Like the ELL model, the residual estimate was also significant (Estimate = 135.88, \( p < .05 \)) and was an indicator (i.e., when compared to the unconditional or prior fitting model) that the model was appropriately specified (see Table 13).

Table 12

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>SE</th>
<th>T</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>43.02</td>
<td>0.89</td>
<td>48.09</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Black</td>
<td>-4.52</td>
<td>1.21</td>
<td>-3.72</td>
<td>.0002</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-2.77</td>
<td>1.38</td>
<td>-2.00</td>
<td>.0460</td>
</tr>
<tr>
<td><strong>Slopes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>13.85</td>
<td>0.32</td>
<td>43.73</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Time*Black</td>
<td>-0.69</td>
<td>0.48</td>
<td>-1.44</td>
<td>.1504</td>
</tr>
<tr>
<td>Time*Hispanic</td>
<td>-0.24</td>
<td>0.54</td>
<td>-0.45</td>
<td>.6554</td>
</tr>
</tbody>
</table>

*Note: \( p \) values that are < .05 are statistically significant.*

Table 13

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>556.99</td>
<td>24.53</td>
<td>.0001</td>
</tr>
<tr>
<td>Time</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Covariance</td>
<td>95.71</td>
<td>6.28</td>
<td>.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>135.88</td>
<td>3.54</td>
<td>.0001</td>
</tr>
</tbody>
</table>

*Note: \( p \) values that are < .05 are statistically significant.*
Assumptions. Investigating the robustness to the SLD data growth curve modeling assumptions was also a pre-requisite to constructing the SLD model. A priori inspections relative to normality, linearity, homoscedasticity, and multicollinearity were examined. Relative to the normality of the distribution, skewness (.42) and kurtosis (3.87) values were approximately normal, although positive values were indicative of outliers resulting in a right skewed and leptokurtic (i.e., sharper peaks and longer tails) distribution. The presence of outliers and any consequential violations to normality in mixed modeling are typically mitigated by large data sets. In the present analysis, 4499 observations were utilized from the 1542 student data files (Raudenbush & Byrk, 2002).
The assumption that the relationship between the residuals and the predicted values was linear was satisfied. A visual inspection of the scatter plot of the residuals and the corresponding predicted values indicated a linear form (see Figure 4). Also, Figure 4 provided support that the homoscedasticity assumption was not violated. A re-inspection of the residuals of the scatter plot indicated that the residuals were mostly centered at or near the zero point and were not formed in any particular pattern; thus the absence of heterogeneity was implied. Like the ELL model, the SLD model also was assessed to rule out multicollinearity concerns by running a parallel regression model. In this instance, variance inflation was calculated on the SLD’s time, race/ethnicity predictor variables (i.e., White, Black, and Hispanic). Numeric findings (i.e., VIF ranged between 0 and 1.16) suggested that collinearity was not a concern for the SLD coefficients. In essence, the variances among time and the race/ethnic coefficients did not appear to be over specified and, by consequence, were not strongly correlated when predicting DORF outcomes (Yu, 2011).

![Plot of level-1 residuals vs. predicted values](image)

*Figure 4. SLD scatter plot of residuals and predicted DORF scores.*
A final a priori assessment was conducted requiring the SLD data to be re-sorted so that student data with 4 or more observations were temporarily eliminated. This assessment was conducted to determine if estimates excluding students with greater than 3 observations (possibly retained students) would significantly change the results. An inspection of the re-sorted data revealed very similar initial (Estimate = 43.24) and slope levels (Estimate = 14.08) estimates.

Summary of Results

This chapter reported findings on two research questions associated with diverse learners and reading trajectories. First, the chapter examined the question: To what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (i.e., ELL-LY, ELL-LF, and ELL-LZ) and Non-ELLS when holding gender and SES constant? The results revealed initial and slope differences between the reference group (i.e., Non-ELL females who were ineligible for free and or reduced lunch) and the respective ELL subgroups. When compared to the reference group, all ELL subgroups had higher positive initial levels, except the ELL-LY students. Slope estimates indicated that all ELL subgroups had steeper trajectories than the reference group, although the differences were not statistically significant for the ELL-LF and the ELL-LZ groups. Initial levels were negative and significant for the Gender and SES predictors. The Gender slope estimate was negative as well. However, the SES slope was positive and significant over time.

Second, the study examined the extent to which the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differed across race/ethnicity groups. Reading fluency (DORF) findings for students with specific
disabilities and their race/ethnicity subgroups revealed that the DORF intercept for the reference group (i.e., White students) was significantly different from the scores of Hispanic and Black students. There were no significant differences in their trajectories over time. Discussions about these findings and their implications for closing the achievement gap between ELL and SLD sub-groups, and their respective reference groups follow in Chapter 5.
Chapter Five
Discussion

The purpose of this study was to investigate the use of the DIBELS as an authentic reading growth measure with diverse student subgroups over time. Specifically, the present investigation examined initial reading levels and trajectories for English Language Learners (ELLs) and students with specific learning disabilities. The two research questions that were examined asked (1) to what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners enrolled in an ESOL program (ELL-LY), ELLs not enrolled but monitored (ELL-LF), ELLs not enrolled and no longer monitored (ELL-LZ), and non-ELLs when holding gender and SES constant? and (2) to what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnicity groups? The present chapter organizes and discusses the findings for the above questions in four ways. First, a summary and discussion about the reading levels, slopes and overall group relationships of the ELLs and students with learning disabilities are presented. Next, potential implications for educational policy and best practice relative to the future growth of students with SLD and ELLs are presented. Third, the limitations and directions for future research are discussed. Finally, the chapter closes with a summary about the salient points of this investigation.
Response to Research Question 1

Findings and discussions about the following question are presented first: To what extent do the intercepts and slopes of third grade reading trajectories differ between English Language Learners (ELLs) enrolled in an ESOL program (LY), ELLs not enrolled but monitored (LF), ELLs not enrolled and no longer monitored (LZ), and non-ELLs when holding gender and SES constant?

An HLM analysis was constructed to address the research question above. Initial level estimates varied for predictors entered into the ELL model. First, the intercept or Non-ELL group (i.e., Non-ELL females who were ineligible for free and reduced lunch) positively and significantly predicted outcomes on the DORF reading measure. When compared to the Non-ELL group, the ELL-LY (i.e., students who were enrolled in an ELL program) group’s scores negatively and significantly predicted DORF scores. In other words, the estimates indicated that ELL-LY students had lower initial DORF scores when compared to their Non-ELL counterparts. The differences observed between Non-ELLs and ELL-LF students were not significant. In fact, like the Non-ELL intercept, the ELL-LF initial intercept positively predicted DORF scores. Similarly, the ELL-LZ (i.e., students who were no longer in a program and were no longer being monitored) group’s initial intercept also demonstrated a positive relationship with the DORF reading measures that was statistically significant. The positive relationships observed in the ELL-LF and ELL-LZ groups indicated higher initial DORF scores by the end of third grade. These findings are consistent with prior research where ELL students outperform their Non-ELL counterparts either in early elementary or by the end of middle school (Baker et al., 2008; D’Angiulli et al., 2004). Studies reporting higher reading
performance in late elementary or in middle school reported that ELL students either attained English proficiency at later ages or that ELL students migrated to the US in later grades (Baker et al., 2008; D’Angiulli et al., 2004; Wiley & Deno, 2005). A recent study by Gutierrez and Vanderwood (2013) suggested similar findings about the relationship between English proficiency and reading fluency; however this study did not include a comparison group. Specifically, Gutierrez and Vanderwood found a direct relationship with ELL proficiency levels and reading fluency where higher English proficiency was associated with higher fluency on the DORF measure. In the present study, higher English proficiency also was equivalent to higher initial DORF scores for the ELL-LZ, ELL-LF and ELL-LY sub-groups, respectively.

The present investigation also found that slope estimates varied. As expected, time was a significant overall predictor of students’ DORF outcomes. Further, the relationship between time and each ELL group was also positive, although the time relationship was only positive and significant for the ELL-LY group. Despite this, the positive estimates suggested that all ELL sub-groups grew at faster rates than Non ELL students. These findings were also consistent with other studies comprising ELL and Non-ELL slopes (Kieffer, 2011; Wiley & Deno, 2005). Specifically, Kieffer found that fluent ELL kindergarten students’ initial levels mirrored Non-ELLS’ initial reading scores by 1st grade because their rate of progress was higher. The author found similar trends with Non-fluent ELL students, but noted their initial levels mirrored Non-ELL students in middle school. Keiffer summarized his findings by stating that ELL students’ slopes were “substantial across time” (p. 1209). Gutierrez and Vanderwood (2013) reported steep slopes among the ELL sub-groups as well. They reported that the steepest slopes
occurred among the most proficient ELL students. In contrast, the greatest growth in the present study occurred among the least proficient ELL students.

The present investigation observed similarities and differences in the SES and gender predictors as well. Both predictors’ initial levels were negatively and significantly related to DORF scores. For example, third grade ELL males who qualified for free and or reduced lunch (an indicator of poverty) were likely to obtain lower initial DORF scores when compared to ELL females who did not qualify for free and/or reduced lunch. Although caution is encouraged when interpreting the present findings, these results are not surprising given that the negative relationship between SES, males, and reading achievement are well documented in the literature (Aud et al., 2011; Aud et al., 2012; D’Angiulli et al., 2004; Morgan et al., 2011). Consistent with the initial level findings for gender (representing males), there was also a negative and significant slope relationship between gender and time. In effect, the estimates suggested that, in general, time did not positively influence the DORF outcomes for Non-ELL or ELL males. In contrast to a negative gender slope, the SES slope was positively and significantly related to time. Stated differently, the relationship between qualifying for free and/or reduced lunch and the time predictor appeared to positively influence DORF outcomes for Non-ELL and ELL students alike. This finding is not necessarily new to the literature, but is sometimes reported differently. For example, D’Angiulli et al. (2004) reported that SES became less of a factor in word-reading as more school instruction was acquired. Thus, inherent in the results is an interaction with SES and time, among other school variables.

Although the relationships discussed above were not causal in nature, the estimates generated from the growth curve analyses permitted cross tabulations between
ELLs, the comparison group, SES and gender predictors. Thus, various diverse student profiles could be generated providing educators with a risk or intensity of intervention indicator so that reading failure was prevented for these at-risk students. For example, male ELL-LY students who qualified for free and/or reduced lunch appeared to be at greatest risk because this group’s initial DORF level was about 56 words read correct with an average unit slope trajectory of 19 words read correct. Using the DORF’s end of third grade benchmark of 110 words read correct, it is highly probable that this ELL-LY group will need additional support in order to increase their probability of success on a high stakes assessment like the FCAT. When third grade reading fluency expectations are placed in the broader context of reading comprehension (e.g., determining the main idea using supporting details in literature based texts or drawing conclusions from informational texts) ELL-LY students’ challenges are even more exacerbated. The literature surrounding both ELL limited English proficient and ELL English proficient students indicate that their lack of comprehension skills are due, in part, to underdeveloped English vocabulary skills (August, Carlo, Dressler & Snow, 2005; Millet, 2011; National Center for Education Statistics, 2012; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Vocabulary results from the 2009 and 2011 National Assessment of Educational Progress reports revealed that 24% of ELL students scored in the lowest quartile (National Center for Education Statistics, 2012). Thus, if grade level performances are expected of ELLs it is critical for educators to further explore the influence of vocabulary as a predictor on reading achievement. Moreover, any substantive findings need to be integrated within an

Although some of the ELL-LY student profiles fell short of the end of third year benchmark (see Figure 1), educators are cautioned about using DORF screenings to generate at-risk profiles about ELL students for three main reasons. First, the cross tabulations calculated were not based on significance testing; significance testing results were only provided for the three ELL groups, the gender and SES predictors, and the comparison group. Second, growth curve points may be different over time. The third caution is that previous research documents that DORF scores alone are not the best predictors for ELL or other high risk populations (Baker et al., 2008; Buck & Torgesen, 2002).

**Response to Research Question 2**

Findings and discussions about the following research question are presented: To what extent do the intercepts and slopes of third grade reading trajectories of students with specific learning disabilities differ across racial/ethnic groups? Significant differences were found for students with specific learning disabilities based on race/ethnicity. Specifically, White students’ initial DORF scores were higher than Hispanic and Black initial DORF scores. These findings mirrored the race/ethnicity differences noted in the general population and the trend among students with specific learning disabilities (Aud et al., 2011; Aud et al., 2012; Cortiella, 2011; Morgan et al., 2011; U.S. Department of Education 2011b; U.S. Department of Education, 2012; Wagner et al., 2006). While the race/ethnicity differences on initial reading scores were clear, explanations about why these differences occurred were beyond the scope of the
analyses conducted in the present study. However, Wagner et al. (2006) reported that low income, in some places, explained as much as half of the seven to thirteen point difference between White students with SLD and their Black and Hispanic counterparts in reading. Morgan, Farkas and Wu (2011) indicated that being a Black elementary student was negatively associated with lower reading performance throughout the elementary school experience, but that underachievement in reading was mitigated when Black students had high initial math skills.

There were no significant differences for slope estimates between the race/ethnicity groups. Specifically, the findings suggested that White, Hispanic and Black students appeared to have similar rates of progress (i.e., 13 to 14 words read correct) for each unit of time observed. However, it was noted that time was a significant predictor of DORF outcomes. Although the present progress rates were not sufficient in closing the achievement gap for students with learning disabilities, it was noted that thirteen to fourteen words read correct was somewhat typical of the number of words gained over twelve to sixteen weeks of instruction (Gutierrez & Vanderwood, 2013).

The need to utilize DORF slopes to progress monitor students with disabilities is especially critical since there are more than two million students with learning disabilities in US schools; many of whom are instructed in general education for about 80% of the school day (Cortiella, 2011). Therefore, follow up investigations about the quality of instruction and the availability, intensity and integrity of scientifically-based intervention options appear to be the next likely steps to determine if the slopes observed in the present study can be steeper in an effort to reduce the learning gap (Batsche et al., 2006; National Research Center on Learning Disabilities, 2007; Vellutino et al., 2006).
Vellutino et al., (2006) noted that reading growth was correlated with designing “specialized” instruction and interventions to meet the unique needs of students. In their study, analyses of the data revealed that there were some “learning disabled” students who scored in the average ranges on basic word skill assessments when provided with tutoring support. One of the purposes of the Vellutino et al.’s study was to demonstrate that high quality instruction in general education (a key variable in the response to intervention framework) reduced over representation in special education and increased academic outcomes for students.

Although evidenced-based interventions and progress monitoring have led to growth for students with learning disabilities, attention also needs to be given to increasing and accelerating these students’ present levels of performance. When the disparity between the reading performance levels of students with SLD and the respective state approved benchmarks (as noted in Figure 3) is examined, it is clear that there is an urgent need to accelerate the learning trajectories of students with learning disabilities to close the reading gap. Education advocates and the U.S. Department of Education believe that examining the quality of instruction given to students with disabilities is one way to change the progress and performance of these students (IDEA, 2004; Greenwood, Tapia, Abbott, & Walton, 2003; Vellutino, Scanlon, Small, & Fanuele, 2006). For example, the IDEA legislation requires schools to use data to document that a student’s underachievement (especially among students who may have learning disabilities) is not due to lack of appropriate instruction. The universal screening assessment procedures that are part of the MTSS process facilitates this because those data allow schools and teachers to re-examine core curriculum and the quality of instruction if less than eighty
percent of the students are not meeting state approved benchmarks. Also, examining instruction through professional development strategies (e.g., coaching and review of best practices) increases teachers’ skills. For instance, Greenwood, Tapia, Abbott and Walton (2003) found that on-going school-wide implementation of evidenced-based literacy practices in early elementary school led to a sustained use of these strategies resulting in greater support of high-risk students. Specifically, the authors reported that the trajectory for high risk students “…was more linear and accelerating as compared to…” the moderate and low-risk student groups (p. 104).

Recent suggestions outlined in the Blue Print for Reform legislation also provide insight into how increasing the levels of performance for SLD students can be accomplished. This legislation calls for equity reform for ‘challenge schools’. Challenge schools are typically low performing schools that have high diversity and high poverty in the student population. As stated in chapter two, equity efforts require states to ensure that high poverty schools receive comparable funding and are given more flexibility to support disadvantaged students (U.S. Department of Education, 2010a). McLaughlin (2010) suggested that states, local education agencies and educators move toward vertical equity as opposed to horizontal equity. Vertical equity suggests that differing and unequal input (e.g., funding, professional development for teachers, specialized services to students and parents) is needed in challenge schools in order to obtain academic outcomes that are commensurate with schools that are less diverse and more academically successful.

**Potential Implications for Policy and Practice**

Despite some of the cautions noted above, the present investigation indicates that inter- and intra-group differences for ELLs and students with disabilities exist at the
respective initial reading levels and, to some extent, within their growth trajectories over time. Moreover, findings appear to suggest that gender and SES influence diverse students’ learning at these inter- and intra-subgroup levels as well. Thus, educators are encouraged to consider the following potential implications for policies and best practices that may be relevant to increasing the achievement of diverse learners.

At the policy level, this study lends support for the utilization of rigorous formative assessment measures as suggested in NCLB (2002) and the proposed reauthorization of ESEA (U.S. Department of Education, 2002; U.S. Department of Education, 2010a). Specifically, systematic monitoring using authentic measures is justified because they identify student needs throughout the year and, in so doing, better inform instruction and best teaching practice. As discussed earlier, the DIBELS were designed to provide information to facilitate instructional support and to enable students to become successful readers. The developers of the DIBELS further reported that DIBELS was never intended to be used in isolation or for high stakes decisions (e.g., retention), but rather, to be part of a decision-making framework that determines students’ response to instruction (Dynamic Measurement Group, 2007). Another policy related implication supported by this study is the importance of examining growth over time and, as such, the need to assess and evaluate students’ performance, especially diverse students, longitudinally. Schleicher, (2010) argued that one of the flaws of NCLB was that the legislation only required a “single bar” assessment of students that occurred once a year. Relative to this investigation, a single measure would not have found how steep the average slope was for ELL-LZ students, nor would a single measure note the homogeneous, more stagnant slope of the students with learning disabilities.
Thus, the current research suggests that simply collecting data on the reading trajectories of diverse students is not enough. Those data must be analyzed in the context of benchmark goals and then used to promote the implementation of strategies to enhance student progress and performance over time.

Best practice suggests that schools and districts should strongly consider the influence of time as a predictor, and actively reflect on its influence for learners. Essentially, the time effect was different for ELLs when compared to students with SLD, and it appeared to vary when interacting with other correlates. For instance, the present study noted a significant and positive interaction between time and SES over and above the significant, but negative impact of gender. Also, time appeared to impact sub-categories of ELL groups differently. Specifically, the ELL-LY group was the most significantly impacted. In contrast, an inverse interaction was noted in the Gutierrez and Vanderwood (2013) study where the more advanced English proficient students obtained steeper slopes over time. Therefore, consideration should be given to strengthening decision making frameworks among schools in Florida that operationalize and quantify time (i.e., how many weeks, months or quarters) so that systematic and consistent monitoring of diverse learners’ response to instruction and intervention are ongoing (Florida Department of Education, 2009; Shinn, 2002; Vellutino et al., 2006). Response to intervention (RtI) is one such framework. As reported earlier, MTSS or RtI is a systemic process that uses data at strategic points in time to evaluate students’ response to instruction or research-based interventions. Inadequate response leads to strategic or increased intensive tiered academic interventions for students at risk for academic failure (Batsche et al., 2006; Fuchs & Fuchs, 2006).
Another potential implication for practice implied by the results of the present study is the importance of monitoring students’ response to instruction. In essence, monitoring students’ response to instruction is another part of the MTSS decision making framework. As stated previously in this chapter, increasing the quality of instruction through professional development techniques (e.g., coaching and review of best practices) is a critical component for changing the outcomes for challenging schools. The benefit of focusing on instruction quality using MTSS strategies will hopefully bolster reading for all students, but also will accelerate the initial reading levels of the most at-risk students like ELLs and enhance the reading trajectories of students with learning disabilities over a shorter period of time. Previous research indicates that the more quickly English proficiency is realized for ELL-LY students, the greater the likelihood that they will perform at reading levels commensurate with their Non-ELL peers (Baker, 2008; Keiffer, 2011).

Limitations

Although the present investigation sheds additional light on the initial and slope level trajectories of subgroups among ELLs and students with specific learning disabilities, limitations surrounding the internal and external validity of this investigation are acknowledged. Internal validity concerns (e.g., issues related to the participants’ data files) are presented first and then the external validity concerns (e.g., the extent to which results are generalizable across the U.S.) are discussed.

Although HLM was considered an appropriate analysis given the nested nature of the data, it was possible that the ELL and SLD models could have been more precise if additional predictor variables were included in the analyses. Specifically, no data were available about vocabulary measures. Evidence in the literature supports the fact that,
under the right conditions (e.g., explicit instruction), vocabulary increases the prediction of reading outcomes for some ELL students; therefore, including a vocabulary predictor variable may better explain some of the remaining residual variance (August, Carlo, Dressler & Snow, 2005; Buck & Torgesen, 2002; Millet, 2011). Related to the internal precision of the model was the statistical weight (i.e., the amount of variance accounted for by each predictor) of each variable in the study. For example, although SES (as measured by free and/or reduced lunch in Florida schools) was included in the ELL analysis, questions remain about whether the ELL predictor or SES status had a stronger effect on ELL student outcomes. However, even though statistical weight analyses were beyond the scope of the present study, cross tabulations of initial level results (see Table 8) for an ELL-LY student without free and/or reduced lunch and a Non-ELL with free and/or reduced lunch suggest that ELL students are at a disadvantage. This is consistent with other reports suggesting that Spanish only and bilingual Spanish speaking first grade students of middle to high SES were very significantly to significantly below the mean of a receptive vocabulary measure (August et al., 2005). In essence, even when the effects of SES are controlled, moderate to high SES Spanish speaking students remain at-risk for reading failure.

Adding SES as a variable in the SLD research question might have provided more precision to the SLD model and, in so doing, would have allowed a deeper exploration of risk factors associated with students with learning disabilities and the race/ethnicity relationship. However, a priori decisions for constructing the SLD model were based on a confirmatory analysis approach specific to inter-group differences about race/ethnicity. Thus, the SES predictor was not included although negative and significant association of
SES with students with SLD was documented (Cortiella, 2011). Relatively, it might be interesting to observe the effects of the model if time was only utilized as a fixed variable and not also as a random variable. Also, adding a school level variable (e.g., school size or average years of teaching experience in a school) may have increased the precision in the present model.

Another limitation specific to the SLD model involved the extent to which the three districts’ definitions surrounding eligibility for SLD services were consistent. It was noted that District A’s definition of specific learning disabilities was somewhat different from District B and District C (see Table 5). Specifically, the definitions used by District B and C were more consistent with the language in the IDEA legislation. In contrast, District A, in some places, summarized the IDEA legislation with umbrella terms and phrases like “…concomitantly with other handicapping conditions…” and “…extrinsic influences…” A brief review of the Florida exceptional student statistics in 2006 revealed that District A had the least amount of students classified with learning disabilities (27%) when compared to 44% and 40% for Districts B and C, respectively (Florida Department of Education, 2008). It is possible that the difference in identification rates between Districts A, B, and C may be due to the variation in definitions used for SLD identification.

The unique history related to each year’s observations gathered by the three districts might also contribute to internal validity concerns. More specifically, student observations obtained in 2006 might be uniquely different from those obtained in 2007 and 2008, given that the observations were from a different group of third grade students. Connected with the unique experiences surrounding the student data were the unique and
different experiences of the schools and the three districts from which the student observations were derived. For example, it was not known if the reading coaches who were typically responsible for data collection and management remained at the same schools across the three years. Also, it was highly unlikely that the same level and intensity of instruction was delivered with identical fidelity across the three districts and respective schools. These two examples could have effected students’ responses on the DORF measure and, in so doing, influence the results of the present study.

Another concern relative to history as a limitation was the fact that it was impossible to group the SLD data by specific (i.e., math, reading or writing) learning disability because the coding structure in the three districts only identified the presence or absence of a learning disability. Although 80% of all learning disabilities is accounted for by reading (Aud et al., 2012), it is unknown if this fact is the same for the three districts studied. Therefore, it is possible that the intercept and slopes results may have been different if the type of learning disability was known and accounted for in the analyses. Experimental mortality of the data was also a limitation to the internal validity of the study. As noted in the ELL descriptive section of chapter three, more than half of the data from the third assessment for ELL students were missing. Although the missing data were adequately handled by the HLM process, questions about the effect of those missing observations remain unanswered.

The Florida Center for Reading Research acknowledged several threats related to instrument administration. For instance, they reported DORF administration errors like assessment administrators starting and stopping the stopwatch at incorrect times, not reading directions verbatim, and not calculating and transferring scores correctly (Florida
Department of Education, 2009). Because the present study data came from districts that, at the very least, engaged in Reading First practices, these issues were presented as possible confounds. Connected to the administrative limitations were other procedural concerns. First, the amount of data checks conducted for the present research was less than 5%. Percentages at or greater than 10% are often reported in the education literature (Baker et al., 2008).

The longitudinal nature and the large sample size of this study increased the external validity needed to better support researchers’ understanding about the intra- and inter-group relationships of diverse student groups within the context of reading fluency. However, limited generalizability is acknowledged because the three participating districts were all in Florida. Moreover, the majority of the sample was retrieved from one large district in South Florida. Essentially, monitoring the effects on third grade reading achievement in District A’s schools may look different in another part of the Southern US or the US in general.

**Potential Implications for Future Research**

In light of the limitations mentioned above, opportunities for future research are implied. First, future research should focus on strengthening external validity by replicating this study selecting districts from across the US randomly. Second, future studies should consider increasing the precision of the growth curve model by adding other within child (e.g., motivation or vocabulary knowledge) and school predictors (e.g., classroom size, average years of teaching experiencing) to the model. For example, modifying research question two to include motivation (e.g., the number points earned on a grade or school wide reading incentive program) as a predictor variable may provide a
prospective about why Black students’ reading scores are the lowest among their counterparts. Modification of some of the internal features of this study could also be considered. For instance, developing a subgroup within the Black race predictor of research question two to include Black Americans and Black Caribbean Islanders may produce interesting findings for educators. A preliminary study using the large sample of Black students enrolled in District A is a good place to start especially because this district has a large immigrant and Caribbean population. Another modification could be to add vocabulary as a predictor variable to question one while still exploring differences between the ELL subgroups. The purpose of this modification would be to determine the influence of the vocabulary predictor among the ELL subgroups. Previous research has suggested that English vocabulary often did not transfer as well among ELL students because these students were less proficient in English, knew less vocabulary words than their Non-ELL counterparts and, by extension, understood less about the word meaning (August et al., 2005). In addition to extending the findings of the present study, examining ELL sub-groups using a vocabulary predictor would also extend the findings in the Millet (2011) study. Millet’s study found that a significant amount of the variance was attributed to vocabulary for ELL second graders; however, the ELL students were not disaggregated by English proficiency levels.

A fifth consideration could be to conduct a longitudinal study using continuous student observations across grades (i.e., DORF data from second through fourth grade), as opposed to just third grade students across years. Finally, future analyses could explore the influence of state and or districts’ definition of students with disabilities in overall identification rates. As mentioned in the limitations section, Districts B and C reported
more students identified with specific learning disabilities when compared to those students identified by District A in 2006.

Conclusion

The purpose of the present study was to address gaps in the literature surrounding the utility of the DIBELS as a formative measure when capturing the intra- and inter-group reading outcomes of two diverse learner groups. Specifically, initial reading levels and slope trajectories for specific categories of ELL students and students with learning disabilities were evaluated. Findings suggested that the DIBELS was useful as a formative assessment measure in providing critical information about the reading levels and rates of progress for both groups of students. Initial reading levels and growth differences were noted for ELL student groups and, in most cases, these differences were negatively influenced when gender and SES were considered as predictors. When categorized by race/ethnicity, initial reading group differences also were observed for a sample of students with learning disabilities; however, their rate of progress was not significantly different. Thus, there is evidence that the DORF provides useful information about the intra- and inter-group reading performances of ELL and students with learning disabilities. The present study also supported the fact that correlates like gender and SES often negatively impact diverse learners’ ability to perform commensurate with their peers in reading. Future research efforts should continue to focus on identifying predictors and overall best practices that will increase academic outcomes for diverse learners.
References


Florida Department of Education (2003). *Update on One Florida equity in education plan and updates on the recommendations from the One Florida accountability commission*. Tallahassee, FL: Author.


Houston, TX:
Appendices
Appendix A: University of South Florida IRB Approval Letter

November 1, 2012

Zhivago Adderley
School Psychology/College of Education
5307 Tummal Court
Wesley Chapel, FL 33545

RE: Exempt Certification Modification Request
IRB#: 106786
Title: Predicting academic outcomes for third grade students: Examining the reading achievement of diverse students using the diagnostic lens of the Dynamic Indicators of Basic Early Literacy Skills

Dear Mr. Adderley:

On 4/11/2008 it was determined that your project referenced above meets the federal criteria, which exempts it from further IRB oversight.

You have requested the following changes to your research:

1. Change in title of study to: "Predicting academic outcomes for third grade students: Examining the reading achievement of diverse students using the diagnostic lens of the Dynamic Indicators of Basic Early Literacy Skills"
2. Change in study length: New anticipated end date of 12/2012
3. Change in number of participants: Increase in total number of participants to 27,000.

On 10/18/2012 the IRB Chairperson reviewed your request and determined this change does not impact the study’s eligibility for exemption. The study continues to meet Exempt Criteria. Any grants supporting this project must be submitted to the Institutional Review Board for review.

Please note that future changes to this protocol may disqualify it from its current exempt status. It is your responsibility to notify the IRB prior to implementing any changes.

Please reference the above IRB protocol number in all correspondence to the IRB c/o the Division of Research Integrity and Compliance. It is your responsibility to ensure that this research is conducted in a manner consistent with the ethical principles outlined in the Belmont Report and with USF IRB policies and procedures.
Appendix A: (Continued)

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,

[Signature]

John Schinka, PhD, Chairperson
USF Institutional Review Board

Cc: Anna Davis, USF IRB Professional Staff
Appendix B: Sample of a Request for District Data Form

DATA REQUEST FORM

<table>
<thead>
<tr>
<th>Name</th>
<th>Zhivago Adderley</th>
<th>Phone</th>
<th>(813) 779-0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>University of South Florida</td>
<td>Today’s Date</td>
<td>1/9/2011</td>
</tr>
<tr>
<td>Email Address</td>
<td><a href="mailto:zadderle@mail.usf.edu">zadderle@mail.usf.edu</a></td>
<td>Associated Deadline</td>
<td></td>
</tr>
</tbody>
</table>

Description of Request:
I am requesting data from Special County Schools to examine the intercept and slopes of two diverse student groups (ELLs and Students with learning disabilities) on their reading achievement.

Grade Levels for which you are requesting data:
- [ ] K
- [ ] 1
- [ ] 2
- [X] 3
- [ ] 4
- [ ] 5
- [ ] 6
- [ ] 7
- [ ] 8
- [ ] 9
- [ ] 10
- [ ] 11
- [ ] 12

Test Administration for which you are requesting data:
- [X] FCAT SSS
- [ ] FCAT NRT
- [ ] Other: DIBELS data

Subject Areas for which you are requesting data:
- [X] Reading
- [ ] Math
- [ ] Writing
- [ ] Science
- [ ] Other: __________________________

Years for which you are requesting data:
- [X] 05-06
- [X] 06-07
- [X] 07-08

How do you wish to receive your report:
- [X] Via Email
- [X] On CD ROM*
- [ ] Color Copies*

* Data on CD ROM and color print copies with student test scores cannot be sent through the courier in accordance with Bay District Schools Confidentiality policy. Please arrange to pick them up.

Send Data Request Forms to:
Zhivago Adderley ▪ Assessment and Accountability ▪ Special District Schools
Phone 000.000.0000 ▪ Fax 000.000.0000

For Data Processing Use Only:

<table>
<thead>
<tr>
<th>Date Received</th>
<th>Date Completed</th>
<th>Priority Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

153
Appendix C: Confidentiality Agreement Form

Confidentiality & Protection of Data Agreement

I promise to keep the electronic data files associated with Zhivago Adderley’s data analysis and dissertation project protected and confidential. I also promise to return or delete the above mentioned data files once I have completed the data checks.

Signed: _________________________ Date: _________________________
Appendix D: ELL Reliability Check Form

Research Question One: Integrity Checklist

Item line Number: _______________________ 

Place a check if you agree that the data line provides the item based on the section criteria or an X if the data line did not provide the item

Overall Presence check

Check that important data are actually present and have not been missed

☐ School ID

☐ Student ID

☐ School Grade (3)

☐ Write the second ORF score here ________

*if missing write /missing/ in the space above

Consistency in variable check

Check fields to ensure data in original data set correspond with the codes in the prepared dataset

☐ Correct ELL code (LY= 1, LF = 1, LZ = 1 or ZZ = 0) transferred

☐ Correct gender code (Female = 0 or Male = 1) transferred

☐ Correct lunch code (Free/Reduced = Y = 1 or No Lunch = N = 0) transferred
Appendix E: SLD Reliability Check Form

Research Question Two: Integrity Checklist

Item line Number: ______________________

Place a check if you agree that the data line provides the item based on the section criteria or an X if the data line did not provide the item

Overall Presence check

Check that important data are actually present and have not been left out

☐ School ID
☐ Student ID
☐ School Grade (3)
☐ Code /K/ present signifying SLD
☐ Write the second ORF score here ________*
  *if missing write /missing/ on the space above

Consistency in variable check

Check fields to ensure data in these fields corresponds from the original dataset to the prepared dataset

☐ Correct Race Ethnicity (W = 0 , B = 1 0, H = 0 1) Code Transferred
☐ Correct Student ID corresponds to code /K/ in original data and student ID in prepared dataset