EXAMINING GROWTH PATTERNS IN HAWAI‘I’S PUBLIC SCHOOL CHILDREN’S READING SKILL DEVELOPMENT, KINDERGARTEN THROUGH GRADE 3: A MULTILEVEL APPROACH

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

EDUCATION

DECEMBER 2014

By

Hugh Howard Dunn

Dissertation Committee

Ronald H. Heck, Chairperson
Morris K. Lai
Peter Leong
Min Liu
Stacey B. Roberts

Acknowledgments

I would like to thank my family for their unconditional love and constant faith in me. I am blessed to have parents who lovingly embrace me and support my every endeavor. I thank my wife, Joanna, and my children Rachel, Karissa, Danae, and Micah who are my constant source of joy and inspiration. They encourage me to be the best that I can be in all areas of my life. Joanna's unselfishness, wisdom, and deep faith in God continue to inspire me. I am also grateful to my best friend, Danny, for always listening, asking the right questions, and keeping my eyes on the target.

In addition, I would like to express my deepest gratitude to the chairperson of my dissertation committee, Dr. Ronald Heck, for his guidance, patience, and mentorship throughout my career. His scholarly accomplishments and passion for research have motivated me to pursue greater academic challenges. I also owe an enormous debt of gratitude to the other members of my committee, Dr. Morris K. Lai, Dr. Peter Leong, Dr. Min Liu, and Dr. Stacey B. Roberts. The wisdom they imparted upon me and the multiple perspectives that they offered helped to advance my thinking and accelerate my growth in the field of educational research and leadership.

Next, I would like to thank my friends and colleagues at the University of Hawaiʻi and the Hawaiʻi Department of Education, especially those affiliated with the Pihana Nā Mamo project, all of whom are too many to mention here. I could not have completed my study without their unrelenting support, timely humor, and words of encouragement.

Most importantly, I would like to thank my Lord Jesus Christ whose daily grace and provision brought this project to fruition.

Abstract

A convergence of evidence supports the notion that a growing number of students enter kindergarten significantly behind their more advantaged peers, and this disparity widens over time. Research demonstrates that the trajectory of reading growth is established early, and performance disparities between strong and weak readers are difficult to close. This early growth pattern is indicative of their future academic success and has important lifelong implications. Thus, early identification of children who are at-risk for reading difficulties is critical to student success.

Statewide assessments lack the analytic power to measure growth and map developmental patterns on foundational reading skills. Thus, reliable measures and methods are essential to identify children's early literacy skill development across primary grades. This study examined Hawai‘i students' developmental trajectories over four years on two Dynamic Indicators of Basic Early Skills (DIBELS) measures, Nonsense Word Fluency (NWF) and Oral Reading Fluency (ORF), designed to monitor acquisition of fundamental reading skills. It examined individual- and school-level factors that may influence students' acquisition of reading fluency, including the extent to which student demographic backgrounds contribute to inter-individual and intra-individual differences in initial status and whether differences increase or decrease in subsequent grades.

The findings indicated that kindergartners demonstrated a wide disparity in initial reading skills, and these gaps persisted over time. Differences due to demographic backgrounds explained variability in students' acquisition of early reading skills more than differences in their school settings. The findings also showed that initial performance predicted rate of growth and that resultant growth patterns, whether cumulative or compensatory, generally differed in terms

of the skill being assessed, student background, and students' initial score on a particular measure.

The study's results imply that there is potential for earlier assessment and intervention in a systematic way to monitor students' literacy development during their initial educational years. The findings also support the utility of longitudinal analysis to examine the ways that students' demographic and literacy backgrounds and their school settings may interact with the consistency and validity of early reading assessments used in classrooms and schools to monitor their acquisition of fundamental reading skills.

Table of Contents

List of Tables

List of Figures

Chapter 1 —Introduction

Introduction to the Problem

There is a convergence of evidence supporting the notion that a growing number of students enter kindergarten significantly behind their more advantaged and typically developing peers, and that over the course of elementary school, the disparity in academic achievement widens (Coyne, Kame‘enui, Simmons, & Harn, 2004; Fiester, 2013; Foster & Miller, 2007). Although research supports that differences in children’s readiness for school in terms of their knowledge and skills as they enter kindergarten can be attributed to variations in family characteristics (e.g., maternal education, family type) and home experiences (e.g., home educational activities, nonparental care), later school success is related to multiple aspects of their development (Bethell et al., 2012), among which the acquisition of early reading skills has been deemed highly critical (Handler & Fierson, 2011; Shaywitz et al., 1994; Stanovich, 1986b; J. K. Torgesen, 2000; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Thus, early identification of children who are at-risk for reading difficulties is essential to preventing and remediating current and future reading problems.

Research has demonstrated that the trajectory of reading growth is established early in children’s school careers and that it becomes increasingly difficult to close the performance disparity between strong and weak readers (Coyne et al., 2004; Good, Simmons, & Kame‘enui, 2001; Juel, 1988; J. K. Torgesen & Burgess, 1998). The early growth pattern of children’s reading skill development is indicative of their future academic success (Lonigan, Burgess, & Anthony, 2000) and has important clinical implications (Logan & Petscher, 2010). Lifelong implications include higher incidences of high school dropout, substance abuse, and behavioral problems (Lyon, 2002; Morgan, Farkas, Tufis, & Sperling, 2008). Furthermore, the level of

reading skills children acquire in elementary school predicts their subsequent educational and occupational pathways (Hock et al., 2009; McNamara, Scissons, & Gutknecth, 2011; Savolainen, Ahonen, Aro, Tolvanen, & Holopainen, 2008; Snow, Burns, & Griffin, 1998; Stanovich, 1986b). Research indicates that as children progress from kindergarten to Grade 3, those who score in the low percentile ranks of reading achievement during these critical years are likely to persist at that level or fall even further behind, and they are more frequently referred for special education services in comparison to their peers who started at higher ranks of reading achievement (Adams, 1990; Mastropieri, Leinart, & Scrugges, 1999; McNamara et al., 2011).

Children who read early and well experience high levels of print exposure and consequent growth in numerous knowledge domains (Cunningham & Stanovich, 1997; Echols, West, Stanovich, & Zehr, 1996; Nese, Park, Alonzo, & Tindal, 2011). In contrast, children who lag behind in their reading skills receive less practice in reading than other children do, miss opportunities to develop reading comprehension strategies (Brown, Palincsar, & Purcell, 1986), often encounter reading material that is too advanced for their skills (M. J. Allington, 1990), and may acquire negative attitudes about reading itself (Oka & Paris, 1986). Stanovich (1986b) referred to this chain of events as the Matthew effect (poor-get-poorer), which results in a steady widening of differences in literacy ability between readers of lesser and greater proficiency, also known as a cumulative trajectory of development (Leppnen, Niemi, Aunola, & Nurmi, 2004).

Student performance trajectories can be extrapolated based on demonstrated performance on particular reading measures across kindergarten and Grade 2; however, most of the literature on the status of children in our nation's schools is focused on subsequent grade levels (e.g., National Assessment of Education Progress main assessments [Grades 4, 8, and 12]; No Child Left Behind Act of 2001 state assessments [grades 3–8 and 10–12]). Nonetheless, it is in the

early elementary grades where the performance disparity between children first appear, and this gap is perhaps the most important policy issue in education in the United States (Slavin, Lake, Chamber, Cheung, & Davis, 2009).

Students' reading performance is a key indicator that educational stakeholders gauge by way of No Child Left Behind (NCLB) state assessments and the National Assessment of Education Progress (NAEP). Although federal legislation aims to "close the achievement gap with accountability" (H. R. 1--107th Congress, 2002, p. 1), there is no national reading test for all students in public schools, particularly any that assess children at multiple points in time across kindergarten to Grade 3. The closest measure is the NAEP reading assessment, which is administered every two years to a nationally representative sample of students in Grades 4 and 8.

Student scores on the NAEP fall within four categories: Below Basic, Basic, Proficient, and Advanced. If the "Below Basic" category is taken to mean students reading below grade level, the performance of Hawaiʻi students on the National Assessment of Educational Progress (NAEP) 2013 validates concern regarding the state of reading proficiency. According to the NAEP, 28% of Hawaiʻi's fourth graders and 29% of its eighth graders read below the basic level (National Center for Education Statistics, 2013b). Moreover, according to the Hawaiʻi state "Bridge" Assessment 2014, which is aligned to the new Common Core standards and designed to transition in 2015 to the new Smarter Balanced assessment, 31% of Hawaiʻi's third grade students are not proficient in reading (Hawaiʻi Department of Education, 2014). These data are informative, but they do not necessarily translate into a systemic strategy to inform reading instruction and subsequently improve student reading achievement.

Statewide assessments are typically administered once near the end of the school year, and thereby lack the analytic power to measure growth across shorter intervals of time (Shapiro,

Solari, & Petscher, 2008) as well as the size of and change in achievement discrepancies among students. Moreover, these data provide a global measure of reading achievement (Hock et al., 2009) and therefore, do not fully uncover the discrepancies in performance on individual factors that are believed to be prerequisite skills that are essential for reading comprehension. It is not sufficient to determine whether a child is "proficient" (i.e., at grade-level) in reading at the end of the year, especially beginning as late as Grade 3. In elementary schools, especially in the early grades, teachers and school leaders also need assessments to give them indicators of potential risk before otherwise undetected weaknesses surface as reading problems (Dickinson & Tabors, 2001; Scarborough, 2002; Snow et al., 1998). Buly and Valencia (2002) asserted that "scores on state tests mask distinctive and multifaceted patterns of students' reading abilities that require different instructional emphases" (p. 219).

Thus, reliable and efficient measures and methods are essential to identify children's early literacy skill development across all primary grades (Fuchs & Fuchs, 1999). Researchers report that the extent to which children acquire foundational skills in two domains of early literacy—code-focused skills (letter knowledge and phonemic awareness) and meaning-focused skills (oral language and comprehension)—shape their future reading development (Ortiz et al., 2012). Furthermore, as cited by Below, Skinner, Fearrington and Sorrell (2010, p. 242), "Research supports that proficient reading is the result of a hierarchical process of skill development" (K. Denton & West, 2002; Johnston, Anderson, & Holligan, 1996; National Reading Panel, 2000; Pugh et al., 2001; Shanahan, 2006). Teachers and school leaders need a way to screen students on these skills and regularly monitor their development before the results show up in test scores that are administered well after student reading deficits should have been identified and appropriate interventions should have been provided.

Unlike states that have adopted statewide reading initiatives such as Ohio (Third Grade Reading Guarantee) Arizona (Move on When Reading), Oklahoma (Reading to Succeed), Mississippi (Literacy-Based Promotion Act) and Colorado (Colorado Read Act), Hawaiʻi has not adopted a statewide, cohesive, and comprehensive policy requiring all public schools, serving kindergarten–Grade 3, to employ assessments designed to address early reading difficulties through ongoing monitoring and timely and targeted intervention. Such a system will capture systematic change early and consistently over time, predict future reading performance, investigate reading development longitudinally, and explain factors that influence variation and patterns in children's early reading development.

Purpose of the Study

This multilevel, longitudinal, retrospective study follows a cohort of students, kindergarten through Grade 3 (2008–2011). The purpose of the study is to explicate and map Hawaiʻi students' developmental trajectory by examining their longitudinal scores on measures designed to monitor early acquisition of fundamental reading skills. It focuses on student background and school context variables that are associated with students' reading proficiency at the beginning of kindergarten and their rate of growth in reading during the first four years of school. DIBELS was chosen as a measure of early reading progress because it has been used in a relatively large subset of public elementary schools in Hawaiʻi to monitor students' early reading skill acquisition years before formal high-stakes state testing is administered.

To accomplish the overarching purpose, methods used in this study are designed to examine more closely students' reading skills, not otherwise detected by the Hawaiʻi State Assessment's (HSA) reform-oriented reading component, to determine the extent to which the HSA outcomes might possibly mask developmental patterns that can be gauged by way of

empirically derived, criterion-referenced scores (Kaminski et al., 2008). A multilevel longitudinal model is employed to examine relationships between student background and school variables in contributing to student growth during their kindergarten–Grade 3 years.

Using a multilevel approach, the study examines potential and meaningful differences in children's growth patterns in reading based on their scores as measured by the Dynamic Indicators of Basic Early Literacy Skills 6th Edition (DIBELS) Nonsense Word Fluency (NWF), Grades K–2, and Oral Reading Fluency (ORF), Grades 1–3, probes. It examines individual- and school-level factors to determine their influence on NWF and ORF growth over time, including whether key predictor variables contribute to inter-individual and intra-individual differences in initial status and if differences increase or decrease in subsequent grades. Additionally, the study examines variances in outcomes in terms of different groups, which will allow the estimation of group averages and group-level effects.

This study extends on disparate findings across studies aimed at examining the extent to which the developmental dynamics of children's reading performance trajectories follow a cumulative growth model, also referred to as the Matthew effect, (Bast & Reitsma, 1998; Leppnen et al., 2004; Onatsu-Arvilommi & Nurmi, 2000; Stanovich, 1986b; Walberg & Tsai, 1983) or a compensatory growth model (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Jordan, Kaplan, Oláh, & Locuniak, 2006). If reading development conforms to a cumulative trajectory, this fan-spread growth path should show stability in individual differences across time, increasing variance, and a positive correlation between the original level of the skill and its subsequent growth (Leppnen et al., 2004). Conversely, if reading development conforms to a compensatory trajectory, this growth path should depict a decrease in individual differences across time and a negative correlation between the initial level of reading performance and its

subsequent growth. In the latter case, however, reading skills may show low, moderate, or even high stability across time. This study takes into account the nonlinear nature of reading growth as well as the effects of student background and school context variables on student growth and achievement (McCoach, O'Connell, Reis, & Levitt, 2006).

Findings will yield critical theoretical and practical implications in the area of policy, programmatic planning, and practice.

Need and Significance of the Study

There is a lack of studies, especially related to Hawai‘i schools, which have employed multilevel longitudinal models to investigate systematically the relationships between student background and school variables in contributing to student growth on early literacy skills using tests such as DIBELS across their kindergarten–Grade 3 years. In addition, there are no studies that examine Hawai‘i DIBELS data across four years to examine whether the developmental dynamics of children's reading performance trajectories on important early reading skills follow a cumulative growth model or a compensatory growth model, and if different growth patterns are associated with specific points in time or groups. Although the core skills targeted by DIBELS have a large body of empirical support underscoring their relevance and utility, this study has the potential to shed light on whether benchmarks that focus on specific measurable skills in early grades, as opposed to broad competencies, are reliable predictors of later reading performance. Moreover, it will make more explicit the developmental connections between early literacy indicators, school data, and reading acquisition.

The results of this study will have specific application to Hawai‘i's public school system, with a research-based reference on the extent to which students' reading skills on fundamental measures vary during their early school years (i.e., kindergarten–Grade 3) and the effects of

child- and school-level covariates of early reading achievement. Information on this level will provide practitioners with informed guidance on programmatic and student-level decisions based on DIBELS performance across years as well as the potential of often overlooked school factors (e.g., enrollment, student composition, educational process, School Quality Survey [SQS] indicators) that may impact children's DIBELS scores. As a result, school systems such as the Hawaiʻi Department of Education should be able to make informed decisions regarding key assets for children's acquisition of basic early literacy skills, effectiveness of support across years, and mobilizing resources to improve programs at the systems level.

Each year, the Hawaiʻi Department of Education (HIDOE) allocates funds to provide all schools with full access to the University of Oregon's Dynamic Indicators of Basic Early Literacy Skills Data System (DIBELS). Although all HIDOE schools have access to this web-based database, the department has not made it mandatory for HIDOE schools to use the system as a formative means to inform instruction and programmatic decisions. Currently, 84 of 207 (40%) Hawaiʻi public schools, serving kindergarten through Grade 5 students, are registered users of the DIBELS Data System (Hawaiʻi Department of Education, 2013; University of Oregon Center on Teaching and Learning, 2014), which gives them consistent access to these measures across grade levels to aid in the early identification of students who are not progressing as expected and provide grade-level feedback toward validated instructional objectives.

Rationale for the Approach to the Problem

Most research involving reading fluency measures of DIBELS (e.g., ORF) has focused on reading performance at a single point in time, but few studies, especially those that have examined reading performance in Hawaiʻi schools, have examined NWF and ORF as a direct index of growth over time. This study will examine children's reading progress, kindergarten

through Grade 3, a period spanning nearly four years before the first formal statewide reading assessment.

The No Child Left Behind Act of 2001 (NCLB) has focused attention on beginning reading in a way that is perhaps unparalleled in recent decades (Invernizzi, Justice, Landrum, & Booker, 2004). Even if future legislation places less emphasis on reading performance as a high stakes indicator, it will not negate or invalidate the large and growing body of research that supports the critical importance of early reading success. As part of NCLB, the *Reading First* initiative was designed to provide assistance to states in the application of scientifically based reading research. Inherent in cohesive scientifically sound reading programs are systematic assessment protocols (Invernizzi et al., 2004) including screening and progress monitoring of children's understanding and application of fundamental component reading skills (phonemic awareness, phonics, fluency, vocabulary, and comprehension). Such data-based approaches to this decision-making process are an important component of evaluating intervention effects, but the degree to which schools, including those in Hawai‘i, actually implement ongoing kindergarten–Grade 3 progress-monitoring using technically adequate data is questionable (Safer & Fleischman, 2005). Without such measures of component skills prior to Grade 3, children's early reading progress and deficiencies are less defined and there is a lack of complete data useful for providing informed instructional interventions. According to Lesaux and Kieffer (2010), reading comprehension is a complex cognitive and linguistic process that should be examined in terms of its relationship to component skills.

Early reading assessment and intervention are supported by the often-observed relation between end-of-first-grade reading performance and subsequent performance (Schwartz, 2005); however, many of the formal measures in our nation's schools emphasize testing in later years.

Moreover, current high-stakes accountability tests do not provide schools with results based on measures that are designed to monitor individual student growth on foundational early literacy skills over time. There is very little information available on kindergarten programs in the United States and on the nation's children as they enter kindergarten and move through the primary grades (West, Denton, & Germino-Hausken, 2000). Kindergarten is a critical period in children's early school careers. For most children, kindergarten represents the first step in a journey through the world of formal schooling. During this time, children's early literacy skills are in a state of gradual maturation (Chaney, 1992; Levine, 1977; Mason, 1980; B. Roberts, 1992; Snow et al., 1998).

According to Invernizzi et al. (2004), wide-scale reading problems can be prevented through early literacy screening at critical developmental junctures, such as entering kindergarten. For example, curriculum-based measurement (CBM) has proved to be a valid and reliable group of assessments that can be quickly administered to students in order to monitor their academic skill development and provide them with appropriate and necessary intervention (Goffreda & DiPerna, 2010). Administering CBM is an evidence-based approach used to measure students' academic proficiency and progress and to evaluate the effectiveness of instruction (Baker et al., 2008; Deno, Fuchs, Marston, & Shinn, 2001).

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is a type of CBM assessment for monitoring student progress on indicators of reading fluency (Baker et al., 2008; Francis et al., 2008; Hall, 2012). Fluency assessments are commonly used as proxies to gauge student reading skill growth because of the theoretical and empirical support for fluency as a key component of reading comprehension and the ease of assessing fluency over comprehension (Baker et al., 2008; Francis et al., 2008). DIBELS measures continue to be one of the most

frequently used progress monitoring assessments in Hawai‘i elementary schools. Within the DIBELS system, benchmarks are administered at least three times annually to all students to identify those who are not meeting critical milestones in early literacy skills. Aggregation of DIBELS data provides systems- and student-level information that are useful to determine the probability of students achieving subsequent early literacy goals as well as the effectiveness of the current instructional supports for helping students to achieve subsequent goals.

Since 40% of Hawai‘i public schools, serving kindergarten through Grade 5, are registered users of the DIBELS Data System (Hawai‘i Department of Education, 2013; University of Oregon Center on Teaching and Learning, 2014), the Hawai‘i Department of Education is poised to use it as a system-level mechanism for examining the effectiveness of instructional reading supports within a classroom, school, or district. Among Darling-Hammond’s recommendations, as cited in the State of Hawai‘i’s Joint Committee Recommendations Regarding Educator Effectiveness System Design and Implementation report (Joint Commettee of the Hawaii State Teachers Association and State of Hawai‘i Department of Education, 2014), were (a) the use of an early reading measure such as DIBELS as a standardized measure for teachers to include in their “basket of evidence” for measuring student learning and (b) the practice of evaluating student academic growth by administering the same assessment at the beginning and end of the year.

Conceptual Framework

This study follows student progress in reading development beginning in their early years in school. Employing a repeated measures design within each school year allows for a refined analysis of progress including how student background as well as features of their classrooms and schools may contribute within- and between-student

differences in early reading growth. The study will include an exploratory analysis of the several levels of nesting in terms of students within schools. This type of examination is well suited to a multilevel framework for examining student outcomes and school effectiveness research, (Rumberger & Thomas, 2000; Shavelson, McDonnell, Oakes, & Carrey, 1987; Willms, 1994), which portrays student learning as multilevel or nested whereby activities at one level are influenced by those at another level (Barr & Dreeben, 1983; Willms, 1994). The study is guided by models consistent with ecological and developmental systems theories that view the lives of children within multiple settings and tethered by multiple relationships (Aikens & Barbarin, 2008).

Creemers and Kyriakides (2010) have recently specified a multilevel conceptual model that proposes student learning growth results from differences in students (i.e., background, skills), direct interactions with teachers, school context and quality educational processes, and the nature of the policy context above the school. This model is useful in understanding that the inputs to schooling are multilevel and student growth requires placing students within a more complete educational context. They argue that efforts to improve student learning should focus on intervening with teachers directly to improve classroom process and assessing and reshaping the school's instructional context surrounding classrooms. Their multilevel model of schooling processes is consistent with Tagiuri (1968) and Willms (1994), who note that the sociological perspective of schooling comprises four major dimensions including ecology (physical and material resources), milieu (characteristics of students and staff), social system (patterns and rules of operating and interacting), and culture (norms, beliefs, values, and attitudes). Berninger, Dunn, Lin, and Shimada (2004) posited that literacy performance

of children in the classroom is affected by cognitive, environmental, and neuropsychological factors. Aaron, Joshi, Gooden, and Bentum's (2008) component model of reading (CMR) takes the multilevel nature of reading acquisition into account. According to Aaron et al. (2008), the three domains that constitute CMR are the cognitive, psychological, and ecological. The study notes that the cognitive domain comprises two components: word recognition and comprehension; the psychological domain includes the components gender differences, learning styles, teacher expectation, motivation and interest, and locus of control; and the ecological domain includes components such as classroom environment, home environment and culture, parental involvement, dialect, and speaking English as a second language.

Given the complex nature of the data relevant to this study, a multilevel analysis was developed in order to account for sources of variation in reading progress due to the presence of nested repeated reading measures within individual students, between students nested within schools, and between schools (Snijders & Bosker, 2012). Failure to account for such sources of variation can lead to erroneous conclusions in trying to account for differences in reading progress between students and between schools. Sloane (2005) noted that reading interventions are multilevel in nature as students are nested in groups that are within classrooms where instruction is delivered. Snijders and Bosker (2012) suggested that "multilevel statistical models are always called for if we are interested in propositions that connect variables defined at different levels, the micro and macro, and also if a multistage sample design has been employed" (p. 10). Multiple contexts should not be neglected due to the dynamic and contextual nature of student development and outcomes (Aikens & Barbarin, 2008).

Research Questions

This study addresses the following primary research questions:

1. What is the shape of the average trajectory and the variability from the average as measured by DIBELS NWF and ORF?
2. To what extent is there a relationship between students' scores on (a) DIBELS NWF and ORF measures and (b) Grade 3 DIBELS Oral Reading Fluency test and Hawai‘i State Assessment Grade 3 reading comprehension test?
3. What 2011 (Grade 3) school-level factors additionally contribute to the closing or widening of the reading achievement gap as measured by DIBELS?

Chapter 2 —Review of Literature

This chapter, divided into five sections, provides a synthesis of the literature related to this research study, which explores the utility of a multilevel model for growth to determine the extent to which potential differences exist in children's early reading growth patterns and how individual- and school-level factors might influence children's early reading growth over time. The literature review also provides contextual data on early reading in terms of children's initial reading proficiency status at the outset of formal schooling, the complex developmental nature of reading acquisition, the assessment of fundamental literacy skills, and different reading growth models, including the Matthew effect.

The body of the literature review is organized around the last section, Early Reading Growth, which discusses longitudinal research that has identified three broad patterns of reading development. The first section outlines the criteria for inclusion in the literature review and describes the scope of the studies. The second section, Early Reading Difficulties, addresses substantive research pertaining to stable and persistent nature of early reading deficiencies and the subsequent difficulties in remediating such deficits over time. It also discusses the roll of response to intervention (RTI) as an early prevention model. The third section, Reading Development, discusses research on reading as a progression of phases within which precursor and fundamental reading skills need to be acquired through early screening, monitoring, and intervention. The fourth section, Early Reading Assessment, highlights key early literacy domains and aspects of a comprehensive assessment system, useful for screening, progress monitoring, and diagnosis. It also describes the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in terms of its background and growth measures. The last section discusses research on discrepancies in reading growth among student subgroups. It addresses effects of

summer setback and frames different development cycles around three general developmental models that distinguish varying patterns of relationships between initial reading status and subsequent reading gains. Finally, this chapter concludes with a summary of the major findings of the literature review.

Introduction and Criteria for Inclusion in Review

This review of literature is a result of a search of library database services (i.e., Academic Search Premiere, MasterFILE Premiere, Primary Search, Professional Development Collection, Psychology and Behavioral Sciences Collection, Teacher Reference Center, TOPICsearch) via the University of Hawai‘i system’s online catalog and EBSCOhost, which is an online reference system that provides customizable basic and advanced searching capacity supported by Boolean logic, natural language, and subject indexing. The literature review search parameters were scholarly (peer reviewed) journals that were published between 2000 and 2014. Based on the topic of this study and the search parameters, the literature review began with a search for relevant literature associated with the following terms: reading disabilities (5,139 references), early reading (5,582 references), reading assessment (2,667 references), multilevel models as they relate to reading (285 references), achievement gap (5,092 references), reading fluency (3,248), response to intervention, Matthew Effect (299) and DIBELS (452). Not all references that were generated had direct application to the topic of this study. Hence, a more refined search was employed using various combinations of the above terms to identify literature with a direct link to the study which can thereby be integrated and generalized across units, treatments, outcomes, and settings (Randolph, 2009). Based on emerging themes, relevant references were selected and coded to delimit the research problem and identify relevant variables. These constructs were then associatively linked to form the organizational framework of this chapter.

Early Reading Difficulties

Over the past decade, the reauthorization of the Elementary and Secondary Education Act, otherwise known as the No Child Left Behind Act of 2001 (NCLB), created a prevailing sense of urgency around the importance of screening all children in the early grades for reading difficulties as a means to "close the achievement gap with accountability" (H. R. 1--107th Congress, 2002, p. 1). NCLB targeted achievement gaps between advantaged and disadvantaged students. In order to close reading gaps, early identification is viewed as a critical first step to providing children with tailored early intervention and supports to ensure that all students acquire adequate reading skills and proficiency (C. A. Denton, 2012).

Although there are no national reading outcomes data on students in the earliest grades, including kindergarten (Kent, Wanzek, & Al Otaiba, 2012, p. 56), past research supports the view that without early identification and intense intervention, poor readers will persist at reading proficiency levels below their current grade (Juel, 1988; Lentz, 1988; Neuman & Dickinson, 2001; Snow et al., 1998; Stanovich, 1986b; J. K. Torgesen & Burgess, 1998; Whitehurst & Lonigan, 2001) and continue to experience reading difficulties in later grades (Adams, 1990; Mastropieri et al., 1999; McNamara et al., 2011). Other studies confirmed the stable and persistent nature of early reading deficiencies and the subsequent difficulties in remediating such deficits over time (Coyne et al., 2004; Cunningham & Stanovich, 1997; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel, 1988; J. K. Torgesen & Burgess, 1998). Juel's (1988) hallmark and oft cited longitudinal study of struggling readers indicates that 88 percent of students who were poor readers in Grade 1 remained poor readers in Grade 4. Similarly, Spira, Bracken, and Fischel (2005) found that children wo are not performing at grade level upon completion of Grade 1 have significantly lower chances of being on or above grade

level in subsequent grades. Research funded by the National Institute of Child Health and Human Development (2000) found that 75% of children who are denied intervention until nine years of age will continue to experience reading difficulties in subsequent grades (Lyon, 1998).

Early identification of children with reading difficulties requires a system that accurately predicts which children are at risk for reading failure. Previous methods for early identification of children with possible reading deficits, however, lacked a strong theoretical base and yielded delayed and unreliable results (Lyon et al., 2001). One such method was based on whether a child had a substantial discrepancy between his aptitude, typically determined by intelligence quotient (IQ) and his reading performance (Francis & Fletcher, 1996; Gunning, 1998). Although the IQ discrepancy-based method, referred as the "wait- to-fail" or discrepancy method, has been widely applied, its conceptual and measurement issues warrant an alternative method for identifying individuals with reading problems (Francis et al., 1996; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992).

The limitations of previous approaches for identifying children with reading difficulties have led to preventive approaches such as response to intervention (RTI), for which related language was written into law with the 2004 reauthorization of the Individuals with Disabilities Education Act, now known as the Individuals with Disabilities Education Improvement Act. Under this law, student eligibility for learning disability services is no longer contingent solely upon a severe discrepancy between achievement and intellectual ability. Critical to RTI are assessment processes involving screening, instructional intervention, and continual monitoring (Boscardin, Muthén, Francis, & Baker, 2008). These processes, beginning with universal screening, are critical to the RTI framework for the early identification of children at risk for

reading difficulties and providing them with intensive intervention to remediate reading problems (Clemens, Hilt-Panahon, Shapiro, & Yoon, 2012; Vellutino, Scanlon, & Zhang, 2007).

Using test scores or other measures from a single point in time, like the IQ discrepancy-based method, however, does not provide timely, reliable and valid data for identifying children with reading difficulties. An effective RTI model ensures that student progress and rate of growth are monitored at regular intervals using measures that yield valid and reliable data, which can then be used to inform and tailor instruction (Clemens, Shapiro, & Thoemmes, 2011).

Reading Development

Understanding students' reading growth requires an appreciation of the complex nature of reading, including its developmental processes and models of development (McCoach et al., 2006). Research has established that acquiring proficiency in reading is the result of a hierarchical process of skill development (Adams, 1990; K. Denton & West, 2002; Johnston et al., 1996; National Reading Panel, 2000; Pugh et al., 2001; Shanahan, 2006). Many theories of reading posit that the array of skills learned across time and different settings (e.g., home, school) coalesce toward automatic reading in Grades 1–5 (Paris, 2005). Although there are persistent debates, also known as the "reading wars," about the developmental order and relevance of the skills that children need to acquire to become proficient readers, in general, reading researchers have embraced the concept that reading development comprises a progression of phases (Adams, 1990; Beech, 2005; Chall, 1983; Frith, 1985; Gough & Juel, 1991; Jackson & Coltheart, 2001; Marsh, Friedman, Welch, & Desberg, 1981; Morris, Bloodgood, Lomax, & Perney, 2003; Seymour & MacGregor, 1984; Spear-Swerling & Sternberg, 1998).

Adams (1990) described reading as the coordinated interplay of multiple skills, including recognizing individual letters, translating letters into sounds, determining the meaning of a word, and comprehending the text as a whole. For mature, fluent readers, these processes are inseparable (Adams, 1990; van Kleeck, 1998). Fundamentally, reading competence requires an individual to identify words and extract meaning from those words (Carlson, Jenkins, Li, & Brownell, 2013). Word reading competence and reading comprehension skills are two basic components of reading performance (Lerkkanen, Rasku-Puttonen, Aunola, & Nurmi, 2004) common in most general reading theories, including the simple view of reading (Gough & Tunmer, 1986) and the component model of reading (Perfetti, 2007).

Over the years, researchers have introduced a range of theoretical frameworks that focus on at least one of the two basic components of reading development and their relationships. Additionally, in the attempt to explain the complicated linkages between phonologically based decoding or word identification skills and language-based skills, researchers have determined that different skills contribute to emergent literacy at different stages (Carlson et al., 2013). For example, Frith (1985) proposed that reading development occurs in three stages, which comprise visual recognition of words (logographic), recognition of grapheme-phoneme correspondences (alphabetic strategies), and orthographic processing. Ehri and Wilce (1983; 1985) suggested that the transition from prereading to initial reading phases is contingent upon the child's application of letter-sound relationships, marked by more automatic and rapid type of processing. Seymour and Evans (1994, 1999) presented the dual foundation model of literacy, which suggests that letter-sound knowledge competency, a critical early literacy skill, mediates the transition into the foundation literacy stage from which higher levels of reading competence evolves. As cited in Beech (2005), Ehri's (1995) stage model, inspired by Frith's model, was based on the notion that

knowledge of the alphabet and letter-sound relations is highly critical in the process of learning to read and spell. Her model proposed four phases of reading development: pre-alphabetic, partial alphabetic, full alphabetic and consolidated alphabetic (Beech, 2005). Scarborough (2005) posited the theory of a two-path model in reading comprehension, one that includes phonological awareness, word recognition, decoding, and spelling, and another that involves language development. The two-path model was supported by studies that demonstrate how some children with reading disabilities have mainly decoding and word-reading deficits but do not exhibit language impairments, whereas other children might lack reading comprehension skills but perform adequately on tasks primarily related to decoding (Shankweiler et al., 1999; Share & Silva, 1987; Silva, McGee, & Williams, 1985). According to Carlson (2013), reading is more complex than the two-path model, as his findings suggested that language and decoding play both independent and interdependent roles in children's paths to comprehension, which supports connectionist theories of word learning (Beck, Perfetti, & McKeown, 1982; Kellas, Ferraro, & Simpson, 1998; Powell, Plaut, & Funnell, 2006; Wolf & Segal, 1999).

Unfortunately, literacy policies and practices fall short in acknowledging that previous research reveals reading is a complex, life-long developmental process in which some students move through the phases of reading development without difficulty, while others struggle at one stage to another (Cantrell, Almasi, Carter, Rintamaa, & Madden, 2010). Nevertheless, the field is beginning to identify essential precursor skills that children should acquire with respect to identified developmental stages or time periods. Carlson (2013) asserted that effective reading intervention requires early identification of reading difficulties based on the understanding of necessary prerequisite skills and the interrelationships among them. A substantial number of theoretical writings, professional opinions, and best-practice documents identify the essential

academic precursor skills that are particularly related to conventional literacy (National Institute for Literacy, 2008).

In their report of the National Research Council's panel on preventing reading difficulties in young children, Snow, Burns, and Griffin (1998) attributed deficits in oral language, phonological awareness (PA), and alphabet knowledge (AK) to the occurrence of significant reading problems. There are notable differences in preschool-age children's competencies on these important precursor literacy skills (Duncan et al., 2007; T. A. Roberts, 2011; Snow et al., 1998), and several studies demonstrate competencies on these skills remain stable across their preschool and elementary school experience (Anthony & Francis, 2005; Lonigan et al., 2000; Storch & Whitehurst, 2002). Whitehurst and Longman (1998) categorized these emergent literacy domains into two components, outside-in and inside-out. Both outside-in and inside-out components are important aspects of children's early literacy experiences and are related to later reading development (Coyne & Harn, 2006). The extent to which children acquire foundational skills in two domains of early literacy—code-focused skills (letter knowledge and phonemic awareness) and meaning-focused skills (oral language and comprehension)—shape their future reading development (Ortiz et al., 2012).

Outside-in components include a variety of oral language skills shown to contribute to a child's reading ability during the preschool period. Oral language skills comprise semantic (word knowledge, expressive and receptive vocabulary), syntactic (knowledge of word order and grammatical rules), and conceptual knowledge (Bishop & Adams, 1990; Bowey, 1994; Demont & Gombert, 1996; Gillon & Dodd, 1994), as well as narrative discourse (ability to construct an original story and retell a recently heard story) (Klecan-Aker & Caraway, 1997; Perfetti, 1985;

Snyder & Downey, 1991; Vellutino & Scanlon, 1985; Vellutino, Scanlon, & Spearing, 1995; Vellutino, Scanlon, & Tanzman, 1991).

Inside-out components encompass code-based skills such as phonological awareness and knowledge of the alphabetic principle (Coyne & Harn, 2006). Phonological awareness refers to an individual's awareness of sounds in spoken words (Stahl & Murray, 1994) and "implicit and explicit sensitivity to the sublexical structure of oral language" (Pullen & Justice, 2003, p. 88). Reading experts commonly describe the alphabetic principle as the systematic relationship between letters and sounds (Pullen & Justice, 2003) characterized by a person's ability to analyze and synthesize individual phonemes and corresponding letters during the decoding process (Nicholson, 1997).

Although reading difficulties for some students are linked to more general language deficits, the vast majority of reading difficulties can be traced to problems related to code-based or phonological-skill deficiencies (Shaywitz et al., 1994; Stanovich, 1986a; J. K. Torgesen, 2000; Vellutino et al., 2004). Currently, the phonological model is deemed one of the most accepted model for the acquisition of reading ability (Handler & Fierson, 2011). Children with deficits in phonological awareness skills have difficulty conceptualizing the alphabetic principle (Uhry & Shepherd, 1997) and are subsequently unable to develop adequate word recognition and decoding abilities (Pullen & Justice, 2003).

In a synthesis of the literature on reading acquisition, Adams (1990) identified five levels of phonological awareness difficulty. He theorized that the most primitive level consists of having the ability to discern audible sounds of words and recall familiar rhymes. The second level was characterized by tasks involving more focused attention to sound components, which requires the ability to recognize and sort patterns of rhyme and alliteration in words. The third

level included blending and syllable-splitting tasks (e.g., isolating initial phonemes), which requires familiarity with the notion that syllables can be divided into phonemes and further divided into sounds of isolated phonemes. Adams (1990) suggested that although the fourth level of difficulty involves tasks that require full segmentation of component phonemes, the most difficult tasks are encountered at the fifth level, where children are required to manipulate phonemes and regenerate the resultant word or pseudoword.

Phonological awareness is necessary to the development of skilled decoding, but it is not sufficient for acquiring the ability to read words (National Reading Panel, 2000; Stanovich, 1992; Tunmer, Herriman, & Nesdale, 1988). Developing word recognition and decoding skills also requires understanding and using the alphabetic principle (Chard, 1995). In alphabetic languages, reading proficiency is dependent upon this foundational skill, which consists of two fundamental sub skills: (a) knowledge of letter-sound correspondences, and (b) the ability to blend phonemes to read words (Luft Baker, Park, & Baker, 2010). Developing early, successful word reading skills, including understanding connected text, is contingent on acquiring automaticity in the alphabetic principle (Adams, 1990; Harn, Stoolmiller, & Chard, 2008a; Hogaboam & Perfetti, 1975; Stanovich, 2000).

According to Torgesen, Rashotte, and Alexander (2001), good and poor readers in English fundamentally differ in their ability to use letter-sound correspondences to decode words. They asserted that poor readers tend to struggle in reading words accurately and instantly, which indicates an incomplete representation of words in memory. Chard et al. (2008) noted that difficulties with decoding have long-term negative implications on children including diminished opportunities to read increasingly complex text, decreased exposure to words and subsequent vocabulary development, and constraints on the development of reading comprehension.

Research has demonstrated positive correlations and longitudinal continuity between individual differences in these code-related skills and later differences in reading ability (Luft Baker et al., 2010).

The preponderance of early reading research in the last 20 years, such as the landmark findings of Snow et al. (1998) and Whitehurst and Longman (1998), provided the initial framework to identify and more closely examine those key developmental precursors to reading and writing abilities, but the panoply of findings are not predicated on a comprehensive summary of the published literature (National Institute for Literacy, 2008). Prior to the convening of the National Early Literacy Panel (NELP) in 2002, however, there was only a small base of systematic empirical summation of research identifying which early literacy skills predict later conventional literacy (National Institute for Literacy, 2008).

In 2002, the NELP met to conduct a synthesis of the scientific research that involved the "measurement of one or more child skills assessed when children were between birth and five years of age or in kindergarten and the measurement of one or more child outcomes on a conventional literacy skill assessed when children were in kindergarten or older" (National Institute for Literacy, 2008, p. 56). The panel retrieved articles that met key causal comparative criteria and presented outcomes on skills that met the conventional literacy criteria. Conventional literacy skills were defined as those that could be classified within the receptive and expressive domains, symmetric to decoding and encoding. The resulting meta-analysis (National Institute for Literacy, 2008) yielded results for two receptive conventional literacy skills (decoding print [words, nonwords, and fluency and measures] and reading comprehension) and one expressive conventional literacy skill (spelling). The meta-analysis suggested that decoding nonwords (average r = .72), spelling (average r = .60) and invented spelling (average r = .58) were among

the 25 identified predictor variables most highly correlated with the conventional literacy skill of decoding. Several variables typically thought of as representing early literacy development (e.g., alphabet knowledge, writing name, phonemic awareness) had average correlations that could be classified as moderate to strong relationships. For example, measures of oral language had a moderate relationship (average r = .33). Overall, oral language was the weakest predictor of decoding among the predictor variables that yielded at least a moderate relationship. Among the 18 identified variables measured in kindergarten or earlier that had the strongest predictive relationship with children's reading comprehension skills were those related to readiness (average r = .59), concepts about print (average r = .54), and alphabet knowledge (average r = .48). Measures of oral language had an average correlation of .33.

The panel's meta-analysis resonated with Storch and Whitehurst (2002), who asserted that inside-out components, or code-based skills, are foundational for children's early reading development in kindergarten and Grade 1. Moreover, a large body of research has determined that the extent to which children in early grades understand phonemic awareness and alphabetic knowledge is highly predictive of their beginning reading skills (Good, Simmons, et al., 2001; O'Connor & Jenkins, 1999; J. K. Torgesen et al., 1999). Although the NELP report emphasized phonological abilities over oral language abilities in early readers, nearly a decade earlier Catts et al. (1999) contended that phonological abilities, a mediator of the relationship between oral language abilities and reading achievement, were generally aimed at explaining word recognition rather than reading comprehension. Thus, a child's phonological ability appears to play a critical role in the early stages of reading acquisition, while the effects of oral language might make its most significant contributions in the later stages (Storch & Whitehurst, 2002).

After screening over 100,000 studies, the National Reading Panel (NRP) published a report in 2000, citing findings that collectively point to the following five "essential components of reading instruction," (see Part B–Student Reading Skills Improvement Grants, Subpart 1–Reading First, Sec. 1208; for definitions, see pp. 203–205): (a) phonemic awareness, (b) phonics, (c) fluency, (d) vocabulary, and (e) comprehension (E. J. Kame‘enui et al., 2006; Shanahan, 2006). These empirically validated foundational skills related to reading outcomes are commonly referred to as the "big ideas" in beginning reading and are considered to be prerequisite and fundamental to reading success (Good, Gruba, & Kaminski, 2002), as well as key measures for which children's demonstrated understanding differentiate successful from less successful readers (E. J. Kame‘enui & Carnine, 1998; Simmons & Kame‘enui, 1998). According to Paris (2005), phonemic awareness and phonics were constrained skills (learned and mastered quickly) while vocabulary and comprehension were unconstrained skills (learned across a lifetime; never fully mastered). Although he cautions against increased assessment and instruction on constrained skills, he noted that mastery of these skills set the stage for reading development, as they enable the acquisition of a wide range of academic skills (Paris, 2005).

Early Reading Assessment

According to Rouse and Fantuzzo (2006), psychometrically sound assessment instruments with demonstrated reliability and construct and predictive validity should measure children's growth on skills (e.g., phonological awareness and alphabetic skills) across time. Since the inception of NCLB, researchers have made significant progress in the area of assessment for the purpose of accurately and reliably assessing children's early literacy skills (Good & Kaminski, 2003; Good, Simmons, et al., 2001; E. J. Kame‘enui et al., 2006; J. K. Torgesen, 2002; Wagner, Torgesen, & Rashotte, 1999).

An important facet of NCLB is the assessment of beginning reading designed "to provide assistance to state educational agencies and local educational agencies in selecting or administering screening, diagnostic, and classroom based instructional reading assessments" (see Part B–Student Reading Skills Improvement Grants, Subpart 1–Reading First, Sec. 1201. Purposes, p. 178). Against this backdrop, the Partnership for Reading, administered by the National Institute for Literacy (NIFL), constituted the Reading First Assessment Committee (RFAC), whose goal was to guide state and local educational agencies in the selection and use of kindergarten through Grade 3 reading assessment instruments that would meet Reading First requirements (E. J. Kame‘enui et al., 2006). The RFAC, also referred to as the Reading First Academy Assessment Committee or Assessment Committee (U.S. Department of Education, 2006), suggests that a comprehensive school-wide early literacy assessment system should include assessments for which resultant data are used for screening, progress monitoring, diagnosis, or outcome evaluation (Coyne & Harn, 2006; E. J. Kame‘enui et al., 2006).

The primary purpose for early literacy assessment is screening all children at the beginning of the school year to determine which children are at risk for failure, and therefore, likely to need additional instruction or intervention (Coyne & Harn, 2006; E. J. Kame‘enui et al., 2006). A second purpose for early literacy assessment is progress monitoring to estimate rates of children's reading improvement and identify those not making adequate growth toward meeting grade-level reading benchmarks. Progress monitoring assessments are administered a minimum of three times a year, using alternate forms of a test. Diagnosis serves a third purpose for early literacy assessment. Diagnostic assessments measure student's strengths and weaknesses across a range of component literacy skills for which results are used to design or plan subsequent interventions tailored to meet individual literacy needs of students (Coyne & Harn, 2006; E. J.

Kame‘enui et al., 2006). Outcome evaluation is typically conducted at the end of the school year to determine if the collective needs of students are being met. High-stakes assessments required by NCLB are examples, which provide relatively little in terms of individual student abilities on early reading skills other than classifying them as having achieved or not achieved grade-level performance standards (Coyne & Harn, 2006).

Perie, Marion, and Gong (2009) categorize assessments into three categories—formative, interim, and summative. They argue that formative assessment, the most narrow of the three types, is linked directly to the current instructional unit and embedded within the activity. Summative assessments are administered once usually at the end of a semester or school year to evaluate the extent to which students' performance meet a set of content standards. Interim assessments, typically administered several times a year, are designed to yield aggregateable data that informs instructional and curriculum decisions at the classroom, school, and district level to meet student needs (Perie et al., 2009). Based on Perie et al's. (2009) description, DIBELS is a type of interim assessment that allows for "aggregation across students, occasions, or concepts." (p. 6)

Background of DIBELS

An example of a comprehensive assessment system, useful for screening, progress monitoring, and diagnosis are the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The RFAC recommended DIBELS as a quality instrument for assessing children's reading skills, kindergarten through Grade 3 (E. J. Kame‘enui, Francis, Good, O'Connor, & Simmons, 2002). In most schools, standardized tests are not administered to students until near the end of Grade 3 (Paleologos & Brabham, 2011). Hence, DIBELS scores "fill a void and generate data for identifying instructional needs, providing more intensive instruction, and monitoring reading

progress for students who otherwise may do poorly on the standardized tests starting in third grade" (Paleologos & Brabham, 2011, p. 55). During school year 2006–2007, the year prior to the first data collection wave of this study, DIBELS was administered in 13,869 schools serving primary grade children across the nation.

DIBELS are the culmination of a program of research and development incepted in the late 1980's by a research team at the University of Oregon (Hall, 2012). DIBELS measures were originally designed based on Curriculum-Based Measurement (CBM) procedures developed in the 1970s by Deno and his colleagues at the University of Minnesota Institute for Research on Learning Disabilities (Good & Kaminski, n.d.). Like CBM, DIBELS is a form of general outcome measure, comprising brief standardized parallel form measures on indicators of students' progress toward achieving global skills growth over time (Hintze, Christ, & Methe, 2006) that are amenable to change through instruction (E. J. Kame'enui & Carnine, 1998). Whereas reading CBM comprises materials sampled from a given curriculum, DIBELS measures feature standardized content that is not directly related to a particular curriculum (Goffreda & DiPerna, 2010).

The DIBELS assessment battery are a set of standardized, individually administered measures designed to assess three key early literacy domains identified by the National Reading Panel: phonemic awareness, alphabetic understanding, and fluency (Coyne & Harn, 2006; Goffreda & DiPerna, 2010). Paris (2005) argued that skills associated with phonemic awareness and alphabetic understanding are among those categorized as constrained (learned and mastered quickly), and if they are analyzed with the same parametric statistical analyses as skills deemed as unconstrained (learned across a lifetime; never fully mastered), like vocabulary and comprehension, resultant findings will be spurious due to their highly variable and unstable

nature across time. Nevertheless, since 1996, DIBELS has received growing recognition as a valid and reliable measure of foundational reading skills. From 1996 to 2001, the Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD), funded by the U.S. Department of Education, supported the research and development on DIBELS (Good, Gruba, et al., 2002). In 2002, the RFAC determined that the research supporting the DIBELS iteration at that time met the minimum validity and reliability criteria as a tool that aligns with one or more of four assessment purposes (i.e., outcome, screening, diagnosis, and progress monitoring) (Hall, 2012).

Since strong correlations have been established between children's DIBELS scores in their early elementary years and their reading performance in later grades, educators are able to identify children at risk for future reading problems and provide them with appropriate instructional interventions (Abbott, Wills, Miller, & Kaufman, 2012). Subsequent to its forerunner, Curriculum Based Measures of Pre-reading Skills, DIBELS has undergone a number of iterations that have led to its current version, DIBELS Next. The DIBELS data relevant to this study are based on student scores on DIBELS 6th Edition across school year 2008–2011. Any mention of DIBELS in this study herein refers to the DIBELS 6th Edition measures and related assessment protocols.

DIBELS Reading Measures

DIBELS measures, consisting of five core indicators, are administered according to grade appropriate subtests three times per year, typically in the beginning of the year/fall, middle of the year/winter, and end of the year/spring. Each indicator (Initial Sound Fluency [ISF], Letter Naming Fluency [LNF], Phoneme Segmentation Fluency [PSF], Nonsense Word Fluency [NWF], and Oral Reading Fluency [ORF]) are designed to measure a fundamental early literacy

skill (Good & Kaminski, 2002b). Two supplemental measures, Word Use Fluency and Retell Fluency, were added to the DIBELS to measure vocabulary/oral language, and reading comprehension, respectively.

Children's DIBELS fluency scores, based on the number of words they read correctly in one minute (correct words per minute [CWPM]), indicate their reading accuracy and speed specific to the targeted sub skill (Abbott et al., 2012). An increasing number of research-based studies related to reading outcomes of children in elementary school, confirm that CWPM is predictive of student's reading comprehension proficiency (Abbott et al., 2012). DIBELS provides criterion and norm-referenced benchmarks (Logan & Petscher, 2010). Student fluency scores on DIBELS indicators can be compared to empirically derived decision categories respective to their grade level, which allows educators to identify deficient skills of struggling students and adjust instruction accordingly. Students who score within the *Low-risk* performance category have an 80% chance of achieving future proficiency. Those who score within the *Some-risk* performance category have a 50% chance of achieving future proficiency. In contrast, students who score within the *At-risk* performance category have only a 20% chance of achieving future proficiency (University of Oregon Center on Teaching and Learning, 2009).

DIBELS ISF, administered in the fall and winter of kindergarten, measures children's phonemic awareness skills, specifically their ability to recognize and produce the initial sound or group of sounds in a word audibly presented to them (Good & Kaminski, 2002b). Standard ISF administration and scoring procedures recommend 16 ISF items per administration, and the student's score is the number of correct initial sounds per minute (Goffreda & DiPerna, 2010).

DIBELS LNF measures children's knowledge of the alphabetic letters throughout kindergarten and the fall of Grade 1 (Good & Kaminski, 2002b). The student is presented with a

page of randomly ordered uppercase and lowercase letters. The student's score is the total number of correct letters she or he verbally names within one minute.

DIBELS PSF is a phonemic awareness skill indicator that is administered in the winter and spring of kindergarten, as well in the fall, winter, and spring of Grade 1 (Clemens et al., 2012; Good & Kaminski, 2002b). The examiner orally administers the PSF probe comprising a monosyllabic word (e.g., "rat"), which the student must segment into individual phonemes (e.g., "/r/ /a/ /t/"). At the end of a minute, the examiner records the total number of phonemes the child correctly produced (Goffreda & DiPerna, 2010).

DIBELS NWF assesses skills associated with the alphabetic principle, including knowledge of letter-sound correspondence and the ability to blend letters into words comprising letters represented by their most common sounds (Good & Kaminski, 2002b). NWF is administered in the winter (optional) and spring of kindergarten, all three Grade 1 assessment points, and the fall of Grade 2 (Clemens et al., 2012; Goffreda & DiPerna, 2010). The student is presented with a list of simple vowel-consonant (VC) and consonant-vowel-consonant (CVC) pseudoword/non-word sequences (e.g., bot, keb, ril) and directed to provide a response by either pronouncing individual letter sounds or the entire words within a one minute time limit. The student's score is the number of correct letter sounds he produces within one minute, either in isolation, as part of a sound segment, or as part of a whole word (Clemens et al., 2012).

DIBELS ORF is used to assess a child's accuracy and fluency in reading connected text (Good & Kaminski, 2002b), and it is administered in the winter and spring of Grade 1 and during all three test assessment points (fall, winter, spring) across Grades 2–6. For the ORF administration, students read aloud three standardized grade-level passages, and the final score is the median number of words read correctly in one minute.

Although the use of DIBELS measures has been challenged on various fronts, including its widespread application under Reading First and debates about its validity (Goodman, 2006b), DIBELS are one of few empirically validated standardized reading fluency assessments available for large-scale (E. J. Kame‘enui et al., 2006), cost-effective progress monitoring (Rouse, 2006). Since being identified as one of the five components of reading instruction included in the National Reading Panel’s (NRP) report (2000), the construct of reading fluency and its relation to instruction and assessments has spurned debate among teachers, policy makers, and researchers (Rasinski, Blachowicz, & Lems, 2012; Samuels & Farstrup, 2006). Questions about the validity of DIBELS ORF as a progress monitoring measure have been central to this debate (Goodman, 2006a; Pressley, Hilden, & Shankland, 2005); however, adding to the substantial empirical base supporting DIBELS in general, there is a growing body of literature and technical reports supporting the validity of DIBELS ORF as a progress monitoring assessment (Paleologos & Brabham, 2011).

ORF is closely associated with Oral Reading measure (R-CBM), which has been widely used for demonstrating validity in measuring overall reading achievement (Clemens et al., 2012). Reliability for ORF has been reported as 0.96 for test-retest (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009). Across multiple investigations, strong concurrent and predictive validity coefficients were found, including excellent sensitivity and specificity levels (≥80%) of reading subtest proficiency on end-of-year district and state standardized assessments (Barger, 2003; Buck & Torgesen, 2003; Good, Simmons, et al., 2001; Hintze, Ryan, & Stoner, 2003; Shaw & Shaw, 2002; Vander Meer, Lentz, & Stollar, 2005; Wilson, 2005). Studies of Grade 2 ORF passages (Good, Kaminski, Smith, & Bratten, 2001) indicated a median alternate form reliability of 0.95 and concurrent validity with the Test of Oral Reading Fluency (Children’s

Educational Services, 1987) of 0.92–0.96. Although these correlations between ORF and state-developed outcome measures provided strong evidence of ORF as a valid measure to predict students' end of year performance, the research is lacking regarding how well ORF predicts outcomes on more comprehensive measures of reading comprehension (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008); however, Reschly, Busch, Betts, Deno and Long (2009) reported a higher correlation among R-CBM and national tests than with state-specific tests. Additionally, Pressley et al. (2005) suggested that performance on ORF is a better predictor of reading comprehension as measured by the TerraNova reading comprehension assessment (CTB/McGraw-Hill, 2004) than teachers' informal assessment of reading comprehension. Across the available literature, Goffreda and DiPerna (2010) synthesized published reliability and validity evidence for each DIBELS measure. Based on their synthesis of 27 studies that met criteria for inclusion in the review, ORF demonstrated strong reliability and validity evidence.

Although evidence for the remaining DIBELS indicators exhibits greater variability, LNF and NWF demonstrate moderate to high predictive validity. According to Catts, Petscher, Schatschneider, Bridges, and Mendoza (2009), NWF was a good predictor of ORF over a 1.5 year period, and ORF was a fair predictor of reading outcome on the basis of the SAT-10 administered in February of Grade 3. Clemens, Hilt-Panahon, Shapiro, and Yoon (2012) provided a comparison of the ISF, LNF, PSF and NWF measures in kindergarten and Grade 1 on their ability to discriminate between children who would later meet an ORF text reading criterion of 35 words per minute or 30th percentile according to DIBELS national normative sample (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002) at the end of Grade 1. They noted that across time, excluding ORF(on which criterion attainment was based), LNF and NWF (largest effect size) were more accurate than other DIBELS measures in discriminating between students

who would later meet or exceeded the ORF reading criterion and those who would not (Clemens et al., 2012). Cummings, Dewey, Latimer, and Good (2011) indicated that both NWF initial status and subsequent scores across all three Grade 1 assessment occasions predicted ORF outcomes, and the predictive power of that relationship was not affected by any change in directions. NWF reliability with Grade 1 students has been reported as 0.94 for test-retest (Harn, Stoolmiller, & Chard, 2008b). Research also demonstrated a high correlation (r = .72) between preschool or kindergarten non-word decoding and later reading (Lonigan, Schatschneider, & Westberg, 2008).

Early Reading Growth

A growing body of evidence (McNamara et al., 2011; Stanovich, 1986b) has confirmed a disparity between good and poor readers as early as kindergarten that grows wider as students enter Grade 3 and subsequent grades. These differences in children are evident prior to their first day of formal education. Fernal, Marchman, and Weisleder (2013) noted significant disparities between infants from higher- and lower-SES families in both vocabulary learning and language processing efficiency that were already present by 18 months, with a 6-month gap emerging by 24 months. Taken together, these findings imply that many of the 58 percent of Grade 4 students achieving at the basic or below basic performance on the most recent National Assessment of Educational Progress (2013a) are students who were at a disadvantage prior to kindergarten and whose reading growth on fundamental early literacy skills remained slow or negligible in comparison to their peers.

There are two broad approaches to closing the achievement gap. According to the preventive view, children enter school on equal footing and schools provide the necessary supports to prevent disadvantaged children from falling behind (Johnson, 2002).The reparative

perspective, supported by much of the evidence on school readiness, argues that disadvantaged children start school at lower performing levels compared to advantaged children, so the role of the school is to help them catch up to higher performing peers (Shin, Davison, Long, Chan, & Heistad, 2013). In either case, the Response to Intervention (RTI) multi-tier approach is designed to identify struggling students and provide them with appropriate interventions at increasing levels of intensity to accelerate their rate of learning (VanDerheyden, n.d.). The concept of growth or rate of improvement in terms of early reading has been paramount in RTI as a means for determining the degree to which a student is responsive to reading instruction (Batsche, Kavale, & Kovaleski, 2006).

Examining children's rate of improvement on early literacy skills as they relate to future reading comprehension outcomes has important instructional and policy implications (Kim, Petscher, Schatschneider, & Foorman, 2010), especially in terms of efforts to close achievement discrepancies. The achievement gap between subgroups of students has been an enduring issue in American education (Rampey, Dion, & Donahue, 2009). Under NCLB legislation, states were required to disaggregate children across underperforming subgroups (i.e., students from economically disadvantaged backgrounds, students with limited English proficiency, students with disabilities, and students from various racial/ethnic backgrounds) as a means to demonstrate adequate yearly progress and determine if the needs of all children are being met (Hosp, Hosp, & Dole, 2011).

Many studies provide calculations showing the extent to which inequality and discrepancies increase during a calendar year; however, it is critical to be mindful of the growing body of research that attributes summer reading setback as a having an enduring negative effect on student performance, particularly for disadvantaged peers (Alexander, Entwisle, & Olson,

2001; R. L. Allington et al., 2010; Downey, von Hippel, & Beckett, 2004). The effects of summer reading setback, which occurs upon students' return to school after summer vacation with diminished reading skills, differs among children from families of varying socioeconomic groups. Existing research indicates that while reading achievement of poor children typically decreases over summer, the reading achievement for their more economically advantaged peers either remains stable or modestly increases (R. L. Allington & McGill-Franzen, 2003). The varying impact of non-school environments on children from different socio-economic backgrounds is aptly illustrated by Hart and Risley, cited by Downey and colleagues (2004), who observed that welfare supported children had "616 words per hour directed to them, compared to 1,251 for children of the working class and 2,153 for children of professionals" (p. 615). Downey et al. extrapolate these patterns further and argue that a five-and-a-half year old child in a professional family has been exposed to 61 million words, while his peer from a working class family has been exposed to 36 million words and his peer on welfare just 18 million words. After kindergarten, these disparities are exacerbated due to varying non-school experiences in the evenings, on weekends, and during summer.

Heyns (1987) was among the first to articulate the equity implications of a seasonal perspective on learning by highlighting achievement patterns in the summer and providing estimates for the influence of disparate home and neighborhood environments that affect student achievement (Downey et al., 2004). Heyns argued that during the summer, when non-school influences were strongest, the discrepancy between disadvantaged and advantaged children's test scores grew, presumably because of non-school factors. She also pointed out that when school is in session, advantaged and disadvantaged children gain cognitive skills at roughly the same rate (Downey et al., 2004).

Hayes and Grether (1983) indicated that an initial seven-month difference in reading achievement between Grade 2 students in 600 schools, categorized as high- and low-poverty schools, increased to a difference of two years and seven months by the end of Grade 6. The researchers also suggested that although reading achievement gains within the academic year (fall-to-spring comparisons) are similar for both high- and low-poverty schools, the effects of summer vacation (spring-to-fall comparisons) on reading achievement demonstrate a different trend. The differential progress between the two groups during the four summers, across Grade 2 and Grade 6, accounted for approximately 80% of the achievement difference. Hence, Hayes and Grether inferred that the large discrepancy between the two groups in Grade 6 reading achievement can be attributed to summer reading setback as well as the smaller initial achievement differences at the beginning of Grade 2.

Borman and D'Agostino (1996) noted discrepancies in reading gains as reported in several large scale evaluations of a federal Title I remedial reading program. They provided data that gains reported for fall-to-spring are substantially larger than gains reported for spring-to-spring, which suggest that effects of intervention diminished during the summer, and intervention during the regular school year alone is not sufficient to sustain gains. Recognizing how the summer setback phenomenon applied to this situation, Borman and D'Agostino recommended that more Title I funds be used to support eligible students during the summers.

More recently, The National Center for Education Statistics (NCES) conducted data analyses on repeated cross-sectional samples, but cross-sectional comparisons alone cannot adequately explain whether students who perform well at the start of school perform even better at later time points in comparison to those students who start school with fewer resources or whose progress slows down at some point in time (Pfost, Hattie, Dörfler, & Artelt, 2014). There

is a need for longitudinal studies that will determine conditional probabilities relative to the performance of poor and good readers during their early years of school to determine whether interindividual differences that occur early increase or decrease with time (Morgan, Farkas, & Wu, 2011). Most longitudinal work focused on reading growth trajectories (Bast & Reitsma, 1998) excluded those children who initially displayed very low levels of academic proficiency (Morgan et al., 2011).

Entwisle, Alexander, and Olson's (1997) longitudinal study of students from Grade 1 to Grade 5 indicated that during the early elementary school years an achievement gap of less than one year that initially separates the poor from their more advantaged peers can increase to almost three years by the end of Grade 5. As such, disadvantaged students can typically be two or three years behind their counterparts upon entering middle school. Entwisle and Alexander (1992) underscored the importance of studying children during the early part of their educational experience when they are "maximally sensitive to home and school influences" and "cognitive growth rates were higher in the first few grades than they are later on" (p. 73).

Typically, learning outcomes demonstrate systemic growth over time that is either linear or nonlinear (McCoach et al., 2006). Although a linear model is the simplest growth curve model within a multilevel model, growth processes do not typically conform to a linear trajectory (McCoach & Kaniskan, 2010). Reading achievement tends to demonstrate a steeper slope during instructional time points and flatter or declining slopes during non-instructional (e.g., summer) time points (McCoach & Kaniskan, 2010). According to McCoach and Kaniskan (2010), multilevel growth models can allow for change in the rate of change captured in the summer setback depicted in children's growth trajectories.

Previous research indicated that linear growth models might not accurately describe growth in reading across points in time due to factors related to individual differences in reading ability at different points in time (Bast & Reitsma, 1998; Leppnen et al., 2004; Lerkkanen et al., 2004). In other words, individual variation in development in terms of cognition or skill may result in different patterns of reading growth trajectories for individual children (Lerkkanen et al., 2004) and their respective subgroups. Yet there is a scarcity of research about growth trajectories in terms of disparate student outcomes across time relative to different subgroups (Morgan et al., 2011).

Longitudinal research focused on reading development demonstrated that as children grow older, a high inter-individual rank-order is normative (Boland, 1993; Butler, Marsh, Sheppard, & Sheppard, 1983; Cunningham & Stanovich, 1997; Juel, 1988; Phillips, Norris, Osmond, & Maynard, 2002). Research on reading development of an entire population of elementary school children indicated high normative change or growth in reading literacy annually, ranging from about a third to one standard deviation. (Bloom, Hill, Black, & Lipsey, 2008; Hill, Bloom, Black, & Lipsey, 2008). Nevertheless, because growth rates are not equal for all students, it is important to investigate whether gains in learning are systematically related to children's initial levels of proficiency (Pfost et al., 2014).

Pfost, Hattie, Dörfler, & Artelt, (2014) summarized 25 years of empirical findings on the development of early inter-individual differences in reading. Based on their synthesis, they suggested that there is empirical evidence for three broad patterns of development. The first type is a cumulative model (Bast & Reitsma, 1998; Bodovski & Farkas, 2007; Kempe, Eriksson-Gustavsson, & Samuelsson, 2011; Leppnen et al., 2004; Onatsu-Arvilommi & Nurmi, 2000; Stanovich, 1986b; Walberg & Tsai, 1983; Williamson, Appelbaum, & Epanchin, 1991)

characterized by increasing differences in student performance as well as variance between them. This trend, termed the Matthew effect (Stanovich, 1986b; Walberg & Tsai, 1983), results in a widening gap or fan-spread pattern of increasing advantage or disadvantage following initial advantage or disadvantage (Morgan, Farkas, & Hibel, 2008; Salaschek, Zeuch, & Souvignier, 2014). As applied to reading outcomes, "over time, better readers get even better, and poorer readers become relatively poorer" (Bast & Reitsma, 1998, p. 1373). Furthermore, Rigney (2010) suggested that absolute and relative Matthew effects can be distinguished within this positive correlation between initial competency and subsequent development. He posited that an absolute Matthew effect can be distinguished by a developmental pattern in which the students who start with high levels of competence show further positive reading literacy gains, whereas the students who start with low levels of competence show negative gains. Similarly, a relative Matthew effect assumes higher reading literacy gains for better readers, but the poor readers demonstrate gains that are flat or marginal (Rigney, 2010).

In contrast, the second type of developmental pattern is a compensatory effect (Aarnouse & van Leeuwe, 2000; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Phillips et al., 2002) or developmental-lag model (Francis et al., 1996; Parrila et al., 2005; Rourke, 1976), which is defined by a pattern where children who enter school with initially less skills demonstrate over time more rapid growth rates than those who enter school with higher skills (Morgan et al., 2011; Salaschek et al., 2014). Consequently, over time, the magnitude of student inter-individual differences will narrow rather than remain constant or widen. Pfost et al. (2014) concluded that this model can be further categorized into two subtypes. On one hand, students with high pretest results will continue to perform well but to a lower extent than their peers with lower pretest results, resulting in a relative compensation. On the other hand, in contrast to an absolute

Matthew effect, students performing initially at high proficiency might experience a decrease in their growth trajectories over time, whereas students performing initially at low proficiency might experience an increase in their growth trajectories (Pfost et al., 2014).

A third model assumes stable differences in proficiency between low- and high performing children, characterized by no expected differences, neither increasing nor decreasing, over time (Pfost et al., 2014).

Of the 28 articles that met Pfost et al's. (2014) criteria for inclusion, there were 78 results that reported on the development of inter-individual differences in reading. Of the 78 results, 33 (42.3%) demonstrated a decreasing achievement gap (compensatory pattern), 20 (25.6%) demonstrate stable achievement differences, and 18 (23.1%) demonstrated an increasing achievement gap (Matthew effect). Additionally, six (7.7%) results showed a delayed compensation pattern, which is characterized by an initial increasing and subsequent decreasing achievement differences. With a ratio of 0.55, Pfost et al. (2014) concluded more studies report a compensatory developmental cycle (e.g., increasing to decreasing achievement gap) in reading growth than argue for a Matthew effect. Shin et al., (2013) posited that inconsistencies in findings might be due to differences in selected measures and their technical limitations. Based on a chi-square test for deviations from expected frequencies, the expected results between the three most salient developmental patterns (decreasing-stable-increasing achievement differences) were not statistically significant ($\chi 2 = 5.606$, degrees of freedom [*df*] = 2, *ns*), and thereby none were overrepresented (Pfost et al., 2014).

Summary

The review of literature provides several insights about the importance of a understanding the complexity of reading, identifying the essential precursor skills that children need to acquire

during critical developmental periods, assessing children in the early grades across multiple points in time on fundamental reading skills, and monitoring children's rate of growth relative to their peers as well as standardized benchmarks in order to meet the needs of all children. These insights set the stage and provide context for pursuing answers to questions posed in this study, which is centered on investigating the potential differences in children's reading growth patterns and the influence of covariates on reading growth over time.

First, since the inception of NCLB, researchers have made significant progress in the area of assessment for the purpose of accurately and reliably assessing children's early literacy skills. Although standardized tests are not typically administered to students until the end of Grade 3 (Paleologos & Brabham, 2011), DIBELS measures beginning in Grade 3 comprise a battery of standardized measures that that can fill the kindergarten to Grade 2 data voids and inform this study, specifically as it relates to Hawai‘i students' growth trajectories across multiple years as measured by DIBELS. Second, there is a growing body of literature that supports DIBELS as comprising psychometrically sound instruments, including its NWF (Harn et al., 2008b; Lonigan et al., 2008) and ORF measures (Barger, 2003; Buck & Torgesen, 2003; Good, Simmons, et al., 2001; Hintze et al., 2003; Shaw & Shaw, 2002; Vander Meer et al., 2005; Wilson, 2005). Third, given the nature of the study, applying a repeated-measures ANOVA growth model or cross-sectional comparisons to the available data for this study will not be sufficient (Bast & Reitsma, 1998; Leppnen et al., 2004; Lerkkanen et al., 2004; Pfost et al., 2014). Repeated measures ANOVA requires complete data and equal spacing of repeated observations, which makes a multilevel model a better option for this study. The complex nature of the data (monitoring progress over a four-year period with two different DIBELS measures) and likely variability in

student growth rates require analyses that can be used to define the complex hierarchical process of reading development more adequately than previous quantitative analyses.

As such, this study proposes to use a multilevel, longitudinal model to examine potential and meaningful differences in children's growth patterns in reading based on their proficiency level as measured by the Dynamic Indicators of Basic Early Literacy Skills 6th Edition (DIBELS) probes relevant to NWF, Grades K–2, and ORF, Grades 1–3. The applied model will situate the findings within the context of this literature review, particularly as they might relate to the three broad patterns of reading development, including the Matthew effect theory.

Chapter 3 —Method

This chapter describes the methods used in conducting this study. It presents a description of the population, sampling procedures, instrumentation, and variables. Also included is a discussion of the design, longitudinal analytic procedures, and the preliminary analyses of the data.

Study Sample Population

Participants in this study consisted of a cohort of students in Hawaiʻi public schools starting kindergarten during school year 2007–2008. There were 17,037 students in the data file with valid pseudo subject identifiers. Of this set of individuals, there was gender information on 14,994 students (52.3% male and 47.7% female) and race/ethnicity information on 14,652 students (28.3% Hawaiian or part Hawaiian, 19.8% Filipino, 15.4% Caucasian, 8.5% Japanese, 7.1% Pacific Islander, 3.8% Hispanic, 3.3% Chinese, 3.3% African American, and 10.5% other race/ethnicity).

Since not all students who were in the data base had reading achievement information relevant to the purposes of the study, further criteria were developed for participation in the study. Students were included if they were assessed with the DIBELS measures at least once during their educational careers from kindergarten–Grade 3 on each outcome measure. This is important for the assumption of data being missing at random (MAR) for conducting analyses of student growth over time (Hox, 2010).

For NWF, there were 8,163 students with at least one NWF measure. Males comprised 53.5% of the sample, and females comprised 46.5%. Their race/ethnicity distribution was similar to the population (31.7% Hawaiian or part Hawaiian, 18.7% Filipino, 14.1% Caucasian, 8.5% Japanese, 7.6% Pacific Islander, 4.4% Hispanic, 3.6% Chinese, 3.1% African American, and

8.2% other race/ethnicity). They were in 75 different schools in kindergarten that administered DIBELS (NWF) in 2007–2008. At the initial test point, 10.3% were identified as receiving special education services. Many students were also from economically disadvantaged families. On average, 57.5% of the students in the participating schools were in the free or reduced federal lunch program.

For ORF, 6,383 students had at least one score. They were 47.2% female and 52.8% male. There were 10.2% of the students receiving special education services and 57.2% eligible for the federal free/reduced lunch program. Their race/ethnicity composition was similar to the group taking the NWF measure (e.g., 29.7% Hawaiian, 18.4% Filipino,16.1% Caucasian). These students were nested in 82 schools during second grade (the middle of the oral fluency data collection).

Instrumentation and Data

The Data Governance Office, Hawai‘i Department of Education, provided the data for this ex post facto study. The principal instrument used to assess literacy progress was the DIBELS 6th Edition subscales, Nonsense Word Fluency (NWF [an index of understanding of alphabetic principle and phonics]) and Oral Reading Fluency (ORF [an index of understanding of alphabetic principle, phonics, accuracy and fluency, and comprehension). Additionally, the Hawai‘i State Assessment (HSA) was included as an index of general reading outcomes. The data on DIBELS are among the most frequently used measures in Hawai‘i for screening and progress monitoring of children’s development of early literacy skills. DIBELS measures have been used in a reasonable subset of public elementary schools in Hawai‘i and across the nation to monitor students’ early reading skill acquisition. The Hawai‘i State Assessment (HSA) is the state’s high stake summative measure for gauging student progress in the attainment of Hawai‘i

content and performance standards. The HSA Reading subtest is administered to students in Grades 3–8 and 10. The outcome data targeted for use in this study included secondary data archived from the previous administration of DIBELS 6th Edition, across school years 2007–2008 through 2010–2011, and the Grade 3 HSA Reading subtest for school year 2010–2011. Of key interest was the development of student reading fluency in the first four grades of primary education. Repeated measures were used to track children's development on skills associated with NWF and ORF measures. In accordance with DIBELS test administration protocol specific to NWF and ORF subtests, each student was assessed at the beginning-of-the-year (BOY), middle-of-the-year (MOY) and end-of-the-year (EOY) in a given academic year. Although DIBELS had some arguable limitations, including subscale changes over time, the present study used the DIBELS scores because the subtests are change-sensitive and not based on fixed-content regardless of age level (Ding, 2012). According to Ding (2012), change-sensitive subscales can reflect the developmental nature of the literacy skills, which provides critical growth data for this longitudinal study.

DIBELS test developers, Good and Kaminski, intended to control for the difficulty of passages so improvement is attributed to the increase in the acquisition of skills and not just differences in the difficulty of passages. For example, Good and Kaminski (2002a) asserted that in developing the ORF passages, short grade-level appropriate passages were first created and analyzed by the Micro Power & Light readability software to determine the readability of passages. They also used the Spache readability formula to refine and revise the passages, which helped to keep the readability of passages in a target range for each grade. When the Spache readability of the passage was higher or lower than the target range, the number of syllables, frequency and difficulty of words, and sentence length were adjusted accordingly. Once

readabilities were determined for all passages, their estimate of relative readability was calculated (Good & Kaminski, 2002a). The DIBELS authors explain that the passages (29 for Grades 2 and 3; 26 for Grade 1) were then arranged in the order of increasing relative readability and assigned to one of three difficulty categories of which the first 9 (8 in Grade 1) were designated the easiest of the passages and the final 8 (8 in Grade 1) were designated as the most difficult of the passages. Of the passages determined to have the easiest relative readability levels, the middle three passages were selected as the benchmarks passages and randomly assigned to be the first, second, or third benchmark assessments. The same was done for the passages in the other two categories. Thus, each of the three benchmark assessments has a set of three passages representative of the three difficulty categories. Good and Kaminski (2002a) argued that the purpose of this protocol was to ensure that benchmark assessments relative to each grade-level were approximately equivalent to each other so that increases in student scores were due to increases in student skills, as opposed to differences in the relative difficulty of the passages. Although the goal was to keep the readability of the passages close and homogeneous in readability, only they noted that only 30% of the student performance variance was explained by the readability formula (suggesting that variance in performance could also be explained by other factors).

The first wave of NWF scores was collected in 2007, during fall of kindergarten. NWF was administered in kindergarten (MOY and EOY), Grade 1 (BOY, MOY, and EOY), and Grade 2 (BOY). ORF was administered in Grade 1 (MOY and EOY), Grade 2 (BOY, MOY, and EOY), and Grade 3 (BOY, MOY, and EOY). Of all the DIBELS subscales, the two that are routinely administered across three or more grade levels and six or more test occasions are NWF and ORF. NWF is the only DIBELS subscale that shares three of the same test occasions with ORF (Grade

1 [MOY and EOY] and Grade 2 [BOY]). Furthermore, NWF and ORF are the only DIBELS measures administered across three grade levels.

Measures

NWF. NWF is a pseudoword-reading test that is a measure of the alphabetic principle, including letter-sound correspondence (Ding, Richardson, & Schnell, 2013; Fien et al., 2010). Further technical description is included in Chapter 2.

ORF. ORF was administered to measure students' accuracy and fluency with connected text (Fien et al., 2010). Spring ORF scores were analyzed in this study as one of the primary end of- year word-reading outcomes in Grades 1–3 (see Chapter 2 for further description).

HSA. The Hawai‘i State Assessment reading subtest is a criterion-references assessment for which scores are used to determine Adequate Yearly Progress (AYP). For this study, it was used as an index for reading comprehension. The students in this study were administered the HSA in the spring of 2011.

Covariates. Covariates used in the present study included demographic data (gender, poverty, ethnicity, and eligibility for special education status). These data were included in the model to control for differences in parameters (e.g., initial scores) and further explain differences in growth patterns in reading skills over time.

- *Gender*. Gender was identified as males, coded "0" and females, coded "1."
- *Student socioeconomic status (SES).* Poverty status was used as an indicator of socioeconomic status (SES). The SES was measured by indicating whether a student was eligible to receive free or reduced price school lunch as determined by the United States Department of Agriculture, Food and Nutrition Service, income eligibility guidelines.

Students who received free or reduced price lunch were coded “1,” and those who were not eligible were coded “0.”

- *Federal ethnicity*. Federal ethnicity was based on guidance published by the U.S. Department of Education, Federal Register, that specifies standards on the collection and reporting of racial and ethnic data by educational institutions and other grantees (U.S. Department of Education, Office of Special Education Programs, 2007). Category options were limited to seven (American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, and Two or More Races) as reported on the official enrollment count date of the cohort of study’s third grade year, effective school year 2010–2011.
- *Race/ethnicity*. This variable comprises 21 categories that are subcategories of several of the federal ethnicity categories. For example, under this variable, the federal ethnicity category of Native Hawaiian or Other Pacific Islander, is further divided to include additional subcategories (e.g., Micronesian, Filipino, Chinese, Guamanian/Chamorro, Indo-Chinese, Japanese, Korean, Other Asian, Other Pacific Islander, Part-Hawaiian, Samoan, Tongan).
- *Special education status*. Special education status was based on multiple disability categories collapsed into one. Students with disabilities were coded “1” and students without disabilities were coded “0.”

School Variables

Chapter 1 *School code*. This variable provides the unique numeric identifier for all 203 Hawai‘i Department of Education elementary schools. Four other variables (KSchool, G1School, G2School, G3School) provide the respective numeric identifier for the school in which each

student was enrolled on the official enrollment count date across each respective school year, 2007–2008, 2008–2009, 2009–2010, and 2010–2011.

Chapter 2 *School context*. These variables include such indicators as student enrollment, percentage of students on free/reduced lunch, average teacher experience, and average proportion of teachers with five years of experience in the school.

Chapter 3 *School processes*. School process variables include implementation of standards-based curriculum, student support services, leadership, and culture of continuous school improvement as measured by the School Quality Survey, Hawai‘i Department of Education.

Data Analysis

School staff administered the student-level NWF and ORF subscales of DIBELS three times per year. Administration of the DIBELS subscales occurred during a two-week assessment window in the months of September–October, February–March, and April–May during each academic year of 2007–2008 to 2010–2011. Students were also administered the HSA during the April–May assessment window at the end of the Grade 3 academic year (2010–2011). All assessments were administered at students’ school sites in accordance with respective assessment protocols.

As noted previously in this chapter, not all individuals had complete data on the repeated DIBELS measures (i.e., NWF, ORF) relevant to this study. Because listwise, pairwise, mean substitution, and other regression-based approaches lead to biased parameter estimation, this study used maximum likelihood (ML) estimation, which is considered acceptable in the literature for working with longitudinal data with missing values, provided the case can be made that the data are missing at random (Heck, Thomas, & Tabata, 2010; Peugh & Enders, 2004). The ML approach selects coefficients that make the observed values most likely to have occurred (Field,

2009). ML estimation is able to “fill” in missing data, since it treats the parameters being estimated as “unknowns” rather than the data points as missing (Peugh & Enders, 2004). Including individuals with partial information is important in establishing the case for data that are missing at random (MAR), since in typical repeated measures analysis of variance (ANOVA) individuals who are missing any data are deleted entirely from the analysis. In this case, this allowed the retention of the 8,163 individuals who took at least one NWF assessment during kindergarten–Grade 1 and the 6,383 individuals who took at least one ORF assessment during Grades 1–3.

Within- and Between-Subjects Growth Models

This study used multilevel growth modeling to examine the relationships between student background and school variables in contributing to students’ early literacy growth during their kindergarten–Grade 3 years. Based on the nested nature of the data, first a two-level growth model was developed to account for the nesting of the repeated measurements within individual students and the possible effects of between-subjects predictors (e.g., gender, race/ethnicity, SES). Two-level modeling has a number of advantages relative to traditional repeated measures ANOVA techniques , including “more accurate standard errors for parameter estimates” (Fien et al., 2010, p. 639). Another advantage previously mentioned is the ability to retain individuals with partial data, which in SPSS is possible when the data are vertically arranged using one or more lines for each subject depending on the number of repeated measures present for each individual (Hox, 2010). In this study, individual students represent Level 2 of analysis, but it can be extended to include schools as the highest clustering level (Level 3).

Following Raudenbush and Bryk’s (2002) general notation, the level-1 model for NWF and ORF can be specified as follows for individual *i* at time *t*:

$$Y_{ti} = \pi_{0i} + \pi_1 linear_{ti} + \pi_2 quadratic_{ti} + \varepsilon_{ti} \ , \qquad (3.1)$$

where π_{0i} is the true individual status for the first measurement occasion, π_1 is the linear effect of growth between occasions, π_2 is a quadratic effect which allows for any acceleration or deceleration over each growth interval, and ε_{ti} represent errors in predicting each individual's growth trajectory. The quadratic parameter indicates that growth for individuals is not constant over time—that is, there is no constant common slope; instead, it represents the curvature parameter associated with individual growth (Singer & Willet, 2003). Negative coefficients indicate that the initial instantaneous growth effect diminishes over time. Whether positive or negative, larger quadratic coefficients indicate more extreme curvature of the trajectory (Singer & Willet, 2003). At level 2, the following models were defined for the random effects:

$$\begin{aligned} \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} + r_{1i} \\ \pi_{2i} &= \beta_{20}. \end{aligned} \qquad (3.2)$$

In Equation 3.2, the β coefficients represent the fixed-effect parameters describing effect of individual growth over time, while the r_{0i} and r_{1i} are random effects for initial status, and the linear growth parameter, respectively. The quadratic effect was found to not vary randomly across individuals so it was fixed to 0 (Raudenbush & Bryk, 2002). At level 2, subsequently, one or more student background variables can be added to the initial status model:

$$\pi_{0i} = \beta_{00} + \beta_{01} X_{1i} + \cdots + \beta_{0q} X_{qi} + r_{0i} \ , \qquad (3.3)$$

where β_{0q} represents the effect of X_{qi} on the initial status intercept. A similar level-2 model can be defined to examine variation in the linear time effect between individuals:

$$\pi_{1i} = \beta_{10} + \beta_{10} X_{1i} + \cdots + \beta_{1q} X_{qi} + r_{1i} \ , \qquad (3.4)$$

where β_{1q} represents the effect of X_{qi} on the linear slope coefficient. At level 2, the variances for the initial status intercept (*I*) and linear slope (*S*), and their covariance (*IS*) are contained in a 2 x 2 covariance matrix:

$$\begin{bmatrix} \sigma_I^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_S^2 \end{bmatrix}. \tag{3.5}$$

Between-School Models

After examining the two-level models, it is possible to add a between-school level for examining possible school effects on student NWF and ORF outcomes. These models can be formulated by adding a subscript *j* to represent schools in the previous set of Equations (3.1-3.4). Random effects of initial status (β_{00j}) and linear growth (β_{10j}) can be defined as randomly varying across schools at level 3:

$$\begin{aligned} \beta_{00j} &= \gamma_{000} + \gamma_{001}W_{1j} + \cdots + \gamma_{0pq}W_{qj} + u_{0j} \\ \beta_{10j} &= \gamma_{100} + \gamma_{101}W_{1j} + \cdots + \gamma_{1pq}W_{qj} + u_{1j}, \end{aligned} \tag{3.6}$$

where γ represents structural parameters, *W* represents one or more school-level predictors, and u_{0j} and u_{0j} represent random effects, which are contained in a 2 x 2 covariance matrix at level 3 (Raudenbush & Bryk, 2002).

It should be noted that the model could not be extended to include classroom-level effects due to the absence of classroom level predictors. According to Fien et al. (2010), this is not considered to be a limitation of the study because reading instruction frequently takes place in multiple settings within a school, especially with the implementation of an RTI framework. More specifically, the RTI framework employs cross-class grouping strategies, making the classroom unit a weak and nebulous predictor, especially when reading instruction occurs in classrooms that are homogenously grouped.

Model 1: Examining Student Progress in NWF and ORF

The first part of the analysis examined children's growth trajectories on two early reading measures (NWF, kindergarten–Grade 2; ORF, Grade 1–Grade 3) to determine the strength of the correlation between Grade 1 ORF and NWF scores across the first two overlapping test occasions (Grade 1 middle-of -year [MOY] and end-of-year [EOY]). Both models generally followed Eq. 3.1 and 3.2. Students' NWF scores were defined as varying over the five occasions (middle-of-year [MOY] kindergarten through the end-of-year [EOY] Grade 1). Initially, as shown in Figure 3.1, student growth in NWF over time was interrupted by summer vacation flat spots. These are apparent in the Figure 3.1 between time 1 and time 2 (end of kindergarten) and between time 4 and time 5 (end of first grade).

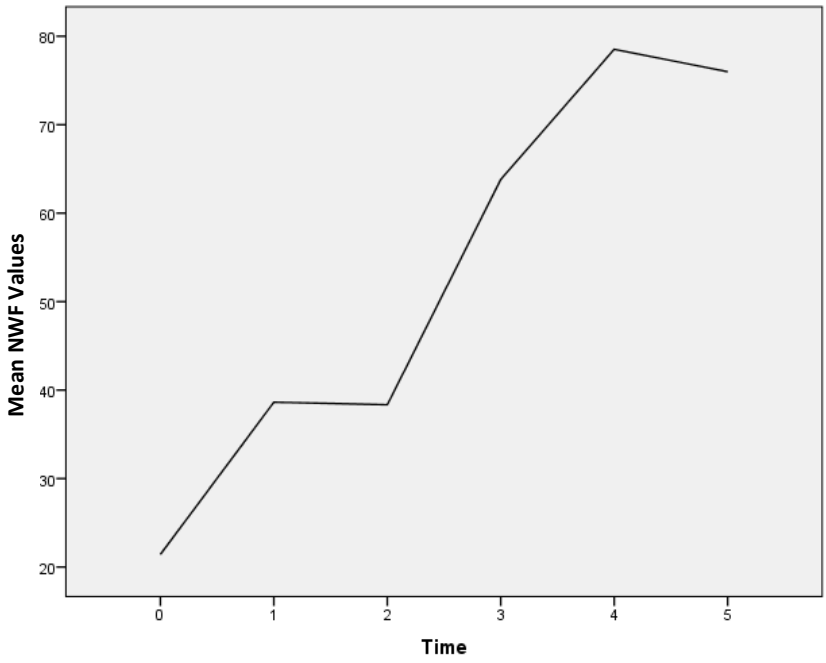

Figure 3.1. Observed growth in NWF scores including summer setback.

One solution to the problem is to code the summer setbacks similar to the previous interval (Raudenbush & Bryk, 2002), which will then produce estimated growth over the period from middle kindergarten to the end of Grade 1, as shown in Figure 3.2. As shown in the figure,

the estimated growth would be from about 20 to about 75 over the kindergarten through Grade 1 repeated measures.

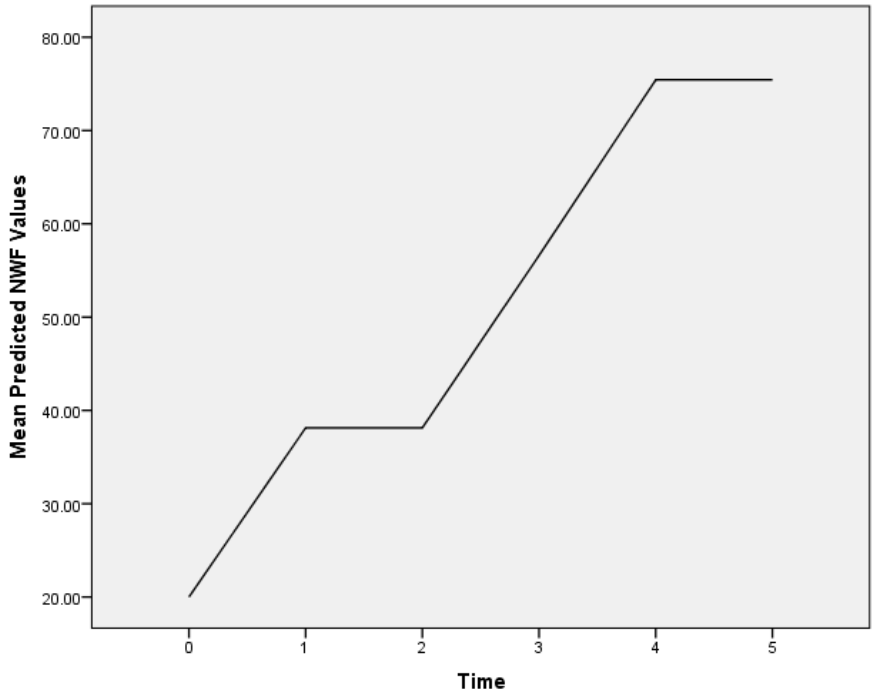

Figure 3.2. Estimated growth in NWF between kindergarten and the end of Grade 1.

A similar situation existed for ORF. Figure 3.3 summarizes the growth in ORF between Grade 1 and Grade 3, again highlighting summer setback periods.

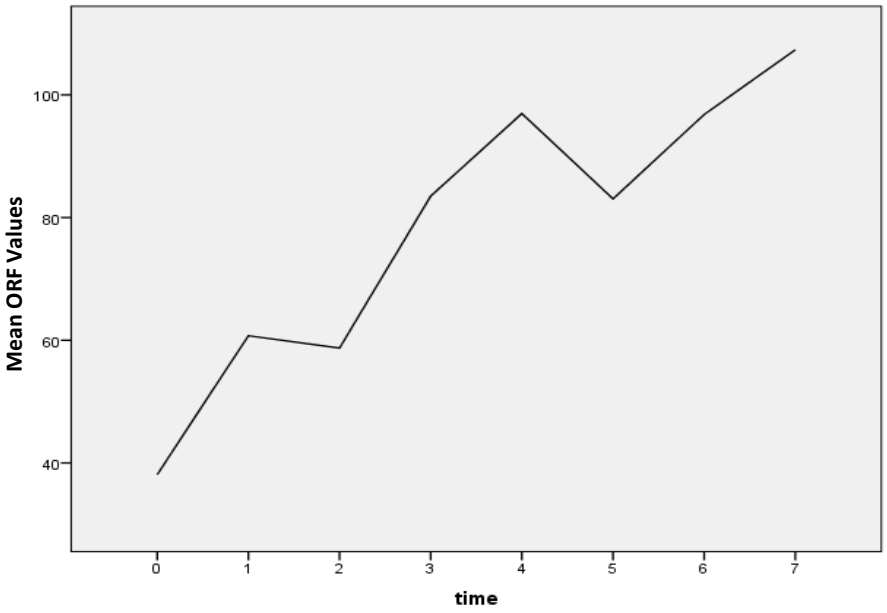

Figure 3.3. Observed growth in oral fluency between Grade 1 and Grade 3.

After applying the similar recoding of the time-related data between time 1 and time 2 and between time 4 and time 5, in Figure 3.4 the estimated growth during this period is summarized.

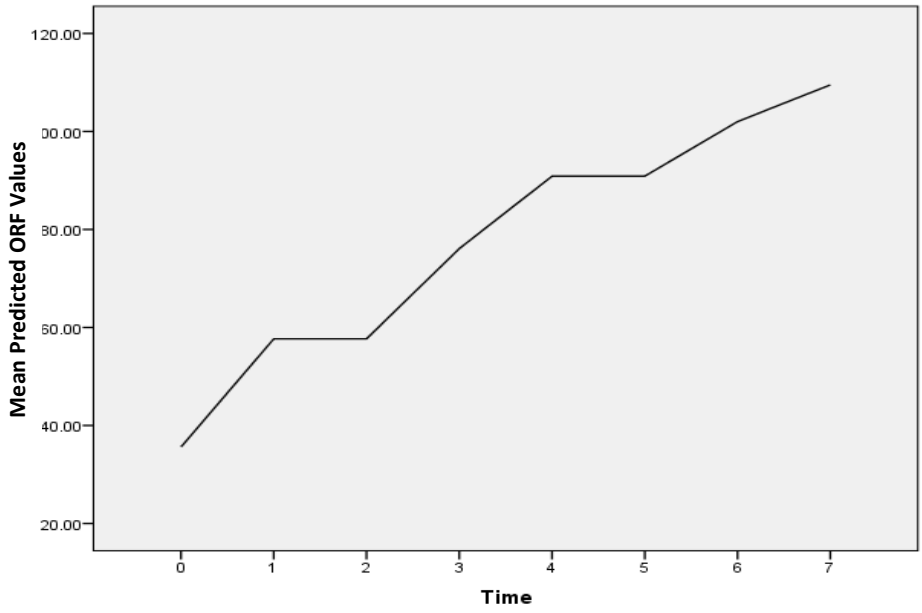

Figure 3.4. Estimated growth in oral fluency between Grade 1 and Grade 3.

Correlations between NWF and ORF measures. Upon establishing the shape and variability in children's growth trajectories on two early reading measures (NWF, kindergarten–Grade 2; ORF, Grade 1–Grade 3), this study analyzed the concurrent correlations between Grade 1 ORF and NWF raw scores across the first two overlapping test occasions (Grade 1 MOY and EOY) and the initial variance components for both measures.

Model 2: Examining Background Variables

Model 2 explored the possible influence of background variables on NWF and ORF growth over time. In initial models a full set of background variables was entered into the model. After all statistically non-significant background variables were removed, final re-estimated models were selected for each outcome that summarize the influence in key background variables on students' reading trajectories. Changes in the model variance components were used to estimate the reduction in variance from the baseline model (Model 1) to the final model with background predictors added (Model 2).

Model 3: Adding the School Variables for Examining Oral Fluency

Finally, a set of school predictors was added to examine whether they provide additional information about how school settings may influence student growth in early reading skills. Model 3 examined how the school settings might also contribute to student learning as measured by ORF. At this stage, preliminary models examined a range of school-level variables (e.g., school context, enrollment, proportion of teachers who had been at the school for five years, average staff teaching experience). One process variable was added that in theory would be related to classroom practices (implementation of standards-based curriculum). School-level variables that were not statistically significant ($p < .05$) were dropped from the final model

presented.

Model 4: Examining the Relationship Between HSA and ORF

Model 4 examined the relationship between the Hawai‘i State Assessment (HSA) Grade 3 scores and the ORF Grade 3 scores. In order to make the comparisons, Model 3 was restructured so that initial status (Grade 1) would correspond most closely with the end of Grade 3 (HSA) score. Five submodels were employed.

The first submodel examined the relationship between the ORF score and the proficiency levels for HSA. This was specified as an ordinal variable with the reference group being category 4 (exceeds proficiency). The second submodel examined the relationship between students' Grade 3 HSA scores expressed as a scaled score (and grand-mean centered) and ORF. In this case, the intercept represents students' final ORF score when the predictors in the model are 0. Submodel 3 was a cross-sectional model that examines the association between grand-mean ORF raw scores and HSA scaled scores at a single point in time (Grade 3 ORF EOY 2001 and Grade 3 HSA Spring 2011. In this case, the intercept represents students' final HSA scaled score when the predictors in the model are 0. Submodel 4 examined the standardized effect of the HSA reading scaled scores on ORF. Since this is a growth model, ORF is the outcome variable. Lastly, submodel 5 investigated the relationship between the grand mean of the last ORF score (Grade 3 EOY) as the predictor and the grand mean of the Grade 3 HSA reading scaled score as the outcome. It also provides an examination of the standardized effect of the ORF on HSA reading scaled scores.

Chapter 4 —Results

The results of this study are presented in this chapter. The two primary aims of the study were (1) to determine the extent to which Hawaiʻi students' reading skills on fundamental measures vary during their early school years (i.e., kindergarten–Grade 3), and (2) to identify child- and school-level covariates of early reading achievement with a particular focus on students who differ in terms of gender, age, socioeconomic status, and ethnicity. The objective of the research was to examine the effects in terms of various groups by including indicators for clusters at multiple levels, which allowed the estimation of group averages and group-level effects. Toward this end, the study employed multilevel longitudinal models to investigate relationships between student background and school variables in contributing to student growth during their kindergarten–Grade 3 years.

Public school students in Hawaiʻi are not formally assessed for accountability purposes (e.g., statewide monitoring of student progress, school compliance with No Child Left Behind Act of 2001) prior to Grade 3. Since HIDOE's only statewide measure for reading performance is the HSA, administered to Grades 3–8 and 10–12 once per year in spring, it does not provide early indicators of potential risk or subtle yet critical changes in children's growth and development of reading skills across multiple points in time within each primary school grade levels. Incremental progress or lack thereof in children's reading skills are masked by attending to high-stakes testing results alone. This study employed methods that can probe beneath the surface of students' scores on the Grade 3 Hawaiʻi State Assessment's reading component to determine the extent to which the HSA scores may hide reading problems that emerge much earlier during children's formal schooling (e.g., kindergarten–Grade 1) and can be adequately assessed by empirically derived, criterion-referenced scores (Kaminski et al., 2008).

Unlike the HSA, DIBELS is designed to measure growth in specified reading skills, not just the status on a skill at a single point in time. Longitudinal analyses of the distribution and fluctuations on skill specific measures across multiple points in time provide more fine-grained information about the reading process especially in terms of reading fluency. Students in the cohort were assessed from Fall 2007 (i.e., the middle of their kindergarten year) to Spring 2011 (Grade 3). The primary assessment tool used to examine individual students' early reading progress at the student level was their (a) Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 6th Edition Grades K–2 measure, NWF, and Grades 1–3 measure, ORF, and their (b) 2011 Grade 3 HSA reading scores. The results help clarify the developmental trajectory for phonemic awareness and fluency with connected text, as measured by NWF and ORF respectively.

The primary parameters of interest in growth analyses are (a) initial status, (b) the rate of increase, and (3) the shape of the growth trajectory (McCoach et al., 2006). This chapter includes these parameters of interest and explains factors that influence variation in student growth in reading. Presented in the chapter are results of four models proposed in this study, organized into four sections.

First, this chapter provides a preliminary description of children's growth trajectories on two early reading measures (NWF, kindergarten–Grade 2; ORF, Grade 1–Grade 3) and describes the strength of the correlation between Grade 1 ORF and NWF scores across the first two overlapping test occasions (Grade 1 middle-of -year [MOY] and end-of-year [EOY]). Second, it then examines a range of background variables and their influence on ORF and NWF growth over time. Third, a set of school predictors is added to examine whether they provide additional

information about how school settings may influence student growth in early reading skills. Finally, the chapter provides an assessment of the relationship between Grade 3 HSA and ORF.

Model 1: Examining Student Progress in NWF and ORF

The following tables and figures are outputs of the initial exploratory analysis of data on Hawai‘i’s ORF and NWF scores across school years 2007–2008 through 2010–2011. A first step in a multilevel analysis of student growth is often to partition the variance in the outcomes into their within-individual and between-individual components.

Nonsense Word Fluency

Table 4.1 provides a summary of the variance components for NWF. The within-individual (repeated measures) variance was 812.00 and the between-individuals variance was 497.27. These estimates represent the grand-mean estimate of students’ NWF as measured over the five occasions (middle-of-year [MOY] kindergarten through the end-of-year [EOY] Grade 1). We can estimate the proportion of variance in NWF that is between students as the ratio of the between-individual variance (497.27) to the total variance (812.00 + 497.27 = 1309.27), or 0.38. This means about 38% of the variance in NWF lies between students and, as we might expect, most of the variance lies within individuals (0.62).

Table 4.1. *Partitioning the Variance in NWF*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Variance	812.001453	7.184419	113.023	.000
Intercept [subject = id]	Variance	497.270281	11.823991	42.056	.000

a. Dependent Variable: NWF (mean = 50.686, SE = 0.302).

Table 4.2 provides an estimation of students’ progress on NWF measured five times from MOY kindergarten through the EOY Grade 1. The table suggests the initial intercept (or students’ true initial status) was almost 20 (19.90), which represented the average score for the

first measurement. Students increased about 18.03 points on average over each successive measurement (i.e., the linear effect), after adjusting for the summer learning flat spots (see Chapter 3 for further discussion). A quadratic effect of time was also added (0.14, $p < .10$), which, because it was positive, suggests there is a slight upward curvature in students' learning progress over time. This suggests that over each interval of study there was a slight acceleration in student learning as measured by NWF. Both the linear and quadratic effects were retained in the final growth model, suggesting a slightly curvilinear shape of the early learning trajectory.

Table 4.2. *Shape of NWF Growth Over Time*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	T	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	19.897909	.233206	6722.572	85.323	.000	19.440752	20.355066
Linear	18.029676	.259595	9884.974	69.453	.000	17.520817	18.538536
Quadratic	.142690	.084893	10964.119	1.681	.093	-.023716	.309095

a. Dependent Variable: NWF.

Table 4.3 presents the variance components associated with students' growth trajectories. Concerning the pattern of repeated measures at level 1 (i.e., the repeated measures nested within individuals), Table 4.3 suggests less variability in students' initial status scores (variance = 108.45) than subsequent measures. More specifically, the table suggests that the individual NWF scores became more varied over time (i.e., variance = 508.11 at time 5, and variance = 470.47 at time 6). The rho estimate is the average correlation between each successive measurement (rho = 0.31), which suggests there was moderate correlation between the successive measurements and provides preliminary evidence that students made considerable progress at increasingly varied rates on this early measure of reading skills during the kindergarten–Grade 1 years. This critical

information, which might be useful in targeting academic resources early in students' elementary years, would be missed when students are not assessed regularly until the end of Grade 3.

Table 4.3 also indicates that the between-student variance [Var(1)] of 251.82, the average initial variance between individuals, was much larger than the initial variance for the repeated measures (108.45). This provides another indication of the considerable amount of variance in reading between students—even in middle kindergarten. A decomposition of the initial variance components within individuals (108.45) and between individuals (251.83) indicated about 43% of the variance in NWF was located between students initially (108.45/360.28 = 0.431). This information provides evidence that differences between students in their early reading readiness was evident from their early entry in kindergarten. The variance for growth between individuals [Var(2)] was statistically significant (46.83 and $p < .001$), suggesting that students' NWF growth rates varied over time. The correlation between the initial status variance and growth variance was statistically significant (0.29, $p < .001$), which suggests a significant relationship between students' initial NWF status and their growth rates. More specifically, because the relationship was positive, it suggests those students who started higher in mid-kindergarten increased this advantage over time, compared with their peers who had lower initial status scores.

Table 4.3. *Variance Components for NWF Model 1*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Var: [Index1=1]	108.451448	6.095138	17.793	.000
	Var: [Index1=2]	196.266293	5.426464	36.168	.000
	Var: [Index1=3]	231.084262	6.270243	36.854	.000
	Var: [Index1=4]	396.532902	11.178828	35.472	.000
	Var: [Index1=5]	508.111001	15.682097	32.401	.000
	Var: [Index1=6]	470.465560	14.969869	31.428	.000
	ARH1 rho	.305035	.013043	23.386	.000
Intercept + linear [subject = id]	Var(1)	251.824771	7.994662	31.499	.000
	Var(2)	46.831678	2.373050	19.735	.000
	Corr(2,1)	.290624	.036111	8.048	.000

a. Dependent Variable: NWF.

Figure 4.1 provides a visual representation of the actual scores for NWF MOY kindergarten to EOY Grade 1) after removing the flat spot in the middle of the trend (between the intervals coded 1 and coded 2, see Chapter 3) associated with the summer three-month period when students were not in school.

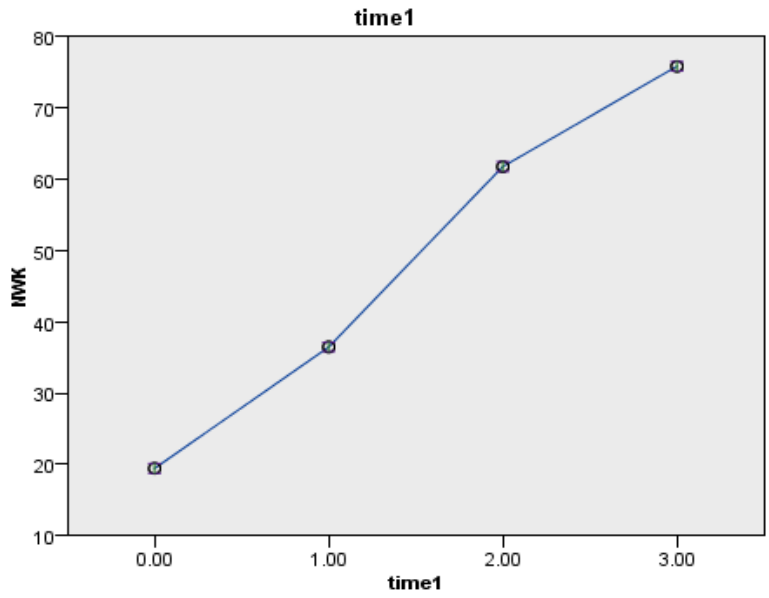

Figure 4.1. Student growth in NWF from MOY kindergarten through MOY Grade 1.

Figure 4.2 suggests that the estimated growth rate was about 18.0 per interval with a slight upward curvature in the estimated growth (see Table 4.2). Hence, adjusting for the flat spots slightly increased the predicted growth rate, resulting in a positive or accelerating curvature.

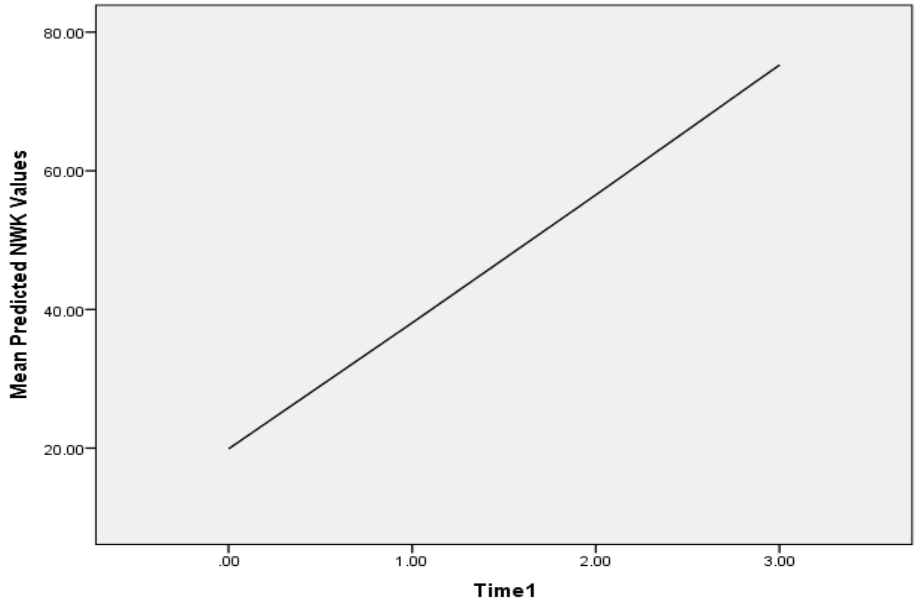

Figure 4.2. Model predicted estimates after adjusting for summer.

Figure 4.3 provides a subset of individuals' actual growth trajectories. The figure helps make the point that although individual trajectories were less varied at the initial status (Time 0), they began to spread out over time, indicating more variance in growth relative to the initial measure. There was also considerable difference in the shape of each individual's actual growth trend over time. This trend explains the weak-to-moderate correlation (rho = 30) of scores between individual and the decreasing predictability of scores across time points.

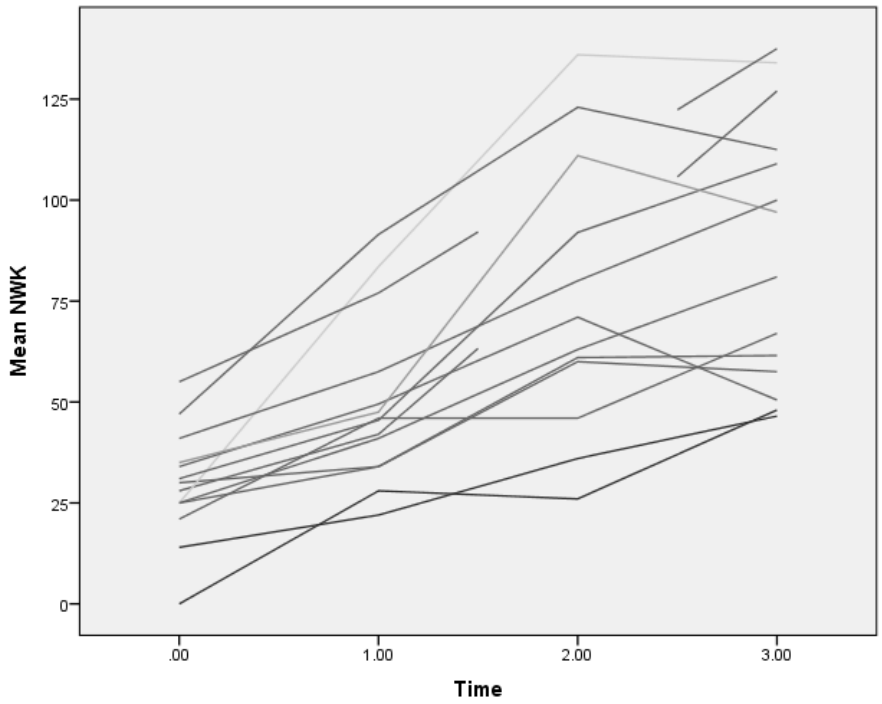

Figure 4.3. Observed NWF scores for a subset of 16 individuals.

Oral Reading Fluency

Table 4.4 provides the initial variance components for ORF. The within-individual component was 683.40 and the between students component was 1019.35. This suggests that about 60% of the variance in ORF was between students (1019.35/1702.75 = 0.599). This is a

much larger proportion of variance between students than observed for the NWF measure (i.e., 0.38).

Table 4.4. *Initial Variance Components for ORF*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Variance	683.398347	5.669526	120.539	.000
Intercept [subject = id]	Variance	1019.352687	21.058738	48.405	.000

a. Dependent Variable: ORF (mean = 74.64. SE = 0.43).

In terms of the trajectory for ORF from middle of Grade 1 to the end of Grade 3, Table 4.5 shows there is considerable growth between occasions that is slightly curvilinear (slowing over time). The average beginning score was almost 36 (35.62). In terms of a linear effect, student ORF scores increased approximately 24 points (23.88) on each successive measurement. Adding a quadratic effect suggested there was a slight slowing of growth over intervals during the study (-1.82, $p < .001$) after accounting for the two summer flat spots. Both the linear and quadratic effects were retained in the model to show the slight curvilinear shape of the early learning trajectory.

Table 4.5. *Growth Trajectories for ORF*

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	Df	t	Sig.	Lower Bound	Upper Bound
Intercept	35.618339	.434605	6178.515	81.956	.000	34.766361	36.470316
Linear	23.882843	.167905	9468.773	142.240	.000	23.553714	24.211973
Quadratic	-1.821131	.029850	6503.361	-61.009	.000	-1.879647	-1.762614

a. Dependent Variable: ORF.

Table 4.6 provides the variance components for ORF. Initially, several variance structures were compared within individuals to find the best fitting structure. Regarding the repeated measure variance for ORF, the variability was more similar over occasions. For

example, the initial variance estimated in MOY Grade 1 (variance = 170.93) was about the same as the variance at time 7 (169.28) and time 8 (101.98). The average correlation between the repeated measures (rho) was 0.33, again indicating a moderate correlation. The variation in growth rate over time [Var(2)] was significant (16.11, $p < .001$). Regarding initial variance, the table suggests there was considerably more variance between individuals (989.54) than within individuals (170.94). A re-estimation of the proportion of variance in ORF between individuals suggested about 85.3% of the variance in ORF was between individuals (989.54/1160.48 = 0.853). This continued the trend observed earlier that differences in individual students' reading became more pronounced during their earlier years in school. Finally, the correlation between the initial status and growth, or the level-2 correlation, was slightly negative (-0.09, $p < .001$), indicating a weak tendency for those starting lower to grow a bit more over time and vice versa.

Table 4.6. *Variance Components for ORF Model 1*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Var: [Index1=1]	170.934952	5.389454	31.717	.000
	Var: [Index1=2]	146.329185	5.639503	25.947	.000
	Var: [Index1=3]	94.792492	3.479236	27.245	.000
	Var: [Index1=4]	184.669192	6.202765	29.772	.000
	Var: [Index1=5]	190.899374	6.020500	31.708	.000
	Var: [Index1=6]	185.020433	5.814759	31.819	.000
	Var: [Index1=7]	169.275547	6.300072	26.869	.000
	Var: [Index1=8]	101.975122	6.806639	14.982	.000
	ARH1 rho	.329136	.015297	21.517	.000
Intercept linear [subject = id]	Var(1)	989.537544	21.228376	46.614	.000
	Var(2)	16.112336	.676860	23.805	.000
	Corr(2,1)	-.094929	.021337	-4.449	.000

a. Dependent Variable: ORF.

Figure 4.4 shows the actual ORF values across all six occasions (MOY Grade 1 to EOY Grade 3) after removing the summer flat spots.

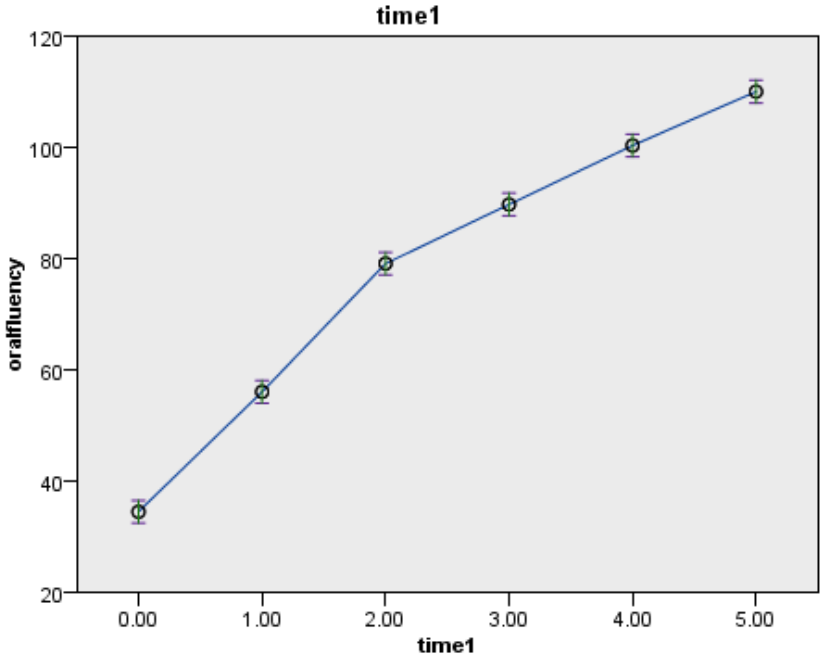

Figure 4.4. Summary of actual ORF scores across occasions after adjustment.

Figure 4.5 shows the predicted values for growth over time after removing the summer vacation flat spots. There figure indicates there was a slight “slowing” in the predicted growth trajectory with the built in flat spots, which adequately models the data.

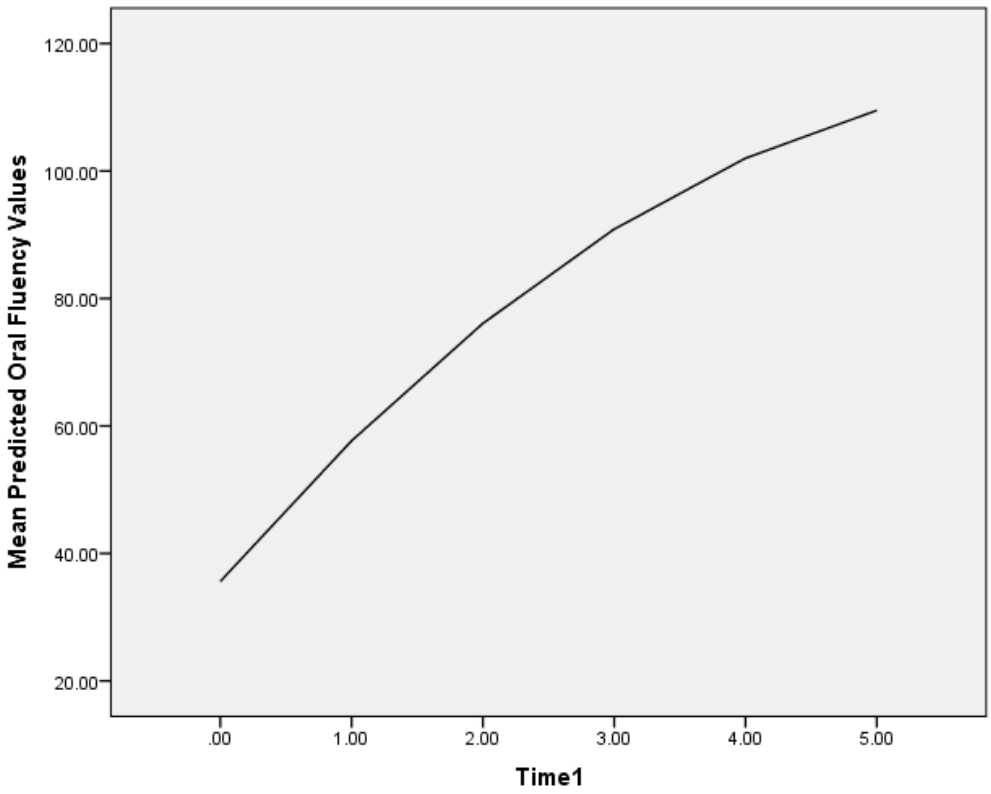

Figure 4.5. Model-predicted student growth in ORF.

Correlations Between NWF and ORF Measures

Having adequately established the shape and variability in children's growth trajectories on two early reading measures (NWF, kindergarten–Grade 2; ORF, Grade 1–Grade 3), a critical next step was to examine the concurrent correlations between Grade 1 ORF and NWF raw scores across the first two overlapping test occasions (Grade 1 MOY and EOY). According to most standards, there was a strong positive correlation observed between the two measures for the first two common occasions of measurement. More specifically, the Grade 1 MOY correlation was 0.74, and the Grade 1 EOY correlation was also 0.74 (see Table 4.7). This suggests a good correspondence between the two measures with an *r*-square estimate (indicating the proportion of overlap) of 0.55.

Table 4.7. *Correlations Between NWF and ORF Measures During Grade 1 (N=5,735)*

Variable	ORF (MOY, Gr. 1)	ORF (EOY, Gr. 1)	NWF (MOY, Gr. 1)	NWF (EOY, Gr. 1)
ORF (MOY, Gr. 1)	1.00			
ORF (EOY, Gr. 1)	0.9*	1.00		
NWF (MOY, Gr. 1)	0.74*	0.72*	1.00	
NWF (EOY, Gr. 1)	0.66*	0.74*	0.72*	1.00

*$p < .001$

Model 2: Examining Background Variables

Model 2 added a range of student background variables in order to determine whether there are differences in initial status and growth rates of subgroups of students. Ideally, if no background variables influenced initial status or growth, it can be assumed that learning is equitably distributed within classrooms and schools and therefore, any observed variability in achievement would be due to what children learn in school. Obviously, the latter cannot be assumed, since students are known to differ in the social capital and associated literacy skills that they bring to school (e.g., where they begin in various instructional sequences) and how quickly they acquire skills and knowledge. For these reasons, background becomes a way of looking at advantages and disadvantages due to factors external to the initial control of the school and teachers. After all statistically non-significant background variables were removed, final re-estimated models were selected for each outcome that summarize the influence in key background variables on students' reading trajectories.

Examining NWF

Table 4.8 summarizes the final fixed-effect estimates for Model 2 after the full range of background variables was added to the model and then the non-significant variables removed, leaving nine groups. The intercept term indicates the expected value of NWF for the reference group (Caucasian) on its first observation. The model suggests students described as disadvantaged (*disadv*), or students who receive free or reduced cost lunch, started 5.85 points lower than their more economically advantaged peers, and females started higher than males (2.48, $p < .001$). The table also identifies various subgroups by race/ethnicity that had higher or lower initial status NWF than the reference group. The initial status among subgroups by race/ethnicity ranged between -3.36 (Micronesians) to 11.96 (Koreans).

Of the nine subgroups in the intercept model, five were retained for their significance in accounting for the variance in students' scores and changes in these scores over time. Regarding growth, disadvantaged students made significantly less growth than their more advantaged peers (-0.57, $p < .05$). There were also a few significant differences in growth related to race/ethnicity that ranged between 1.46 (Micronesian) and 2.96 (Other Asian). Possible quadratic interactions (Disadv*quadtime1) were also investigated but were not found to be statistically significant, so they were not retained in the final NWF model. Overall Table 4.8 suggests considerable social distribution of learning among early primary-grade students.

Table 4.8. *Model 2: NWF Final Background Model*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	21.213258	.525913	5331.770	40.336	.000
Disadv	-5.849014	.533187	5010.286	-10.970	.000
Female	2.476380	.491241	5266.097	5.041	.000
Chinese	10.648163	1.367197	4999.712	7.788	.000
Filipino	5.981837	.648720	4800.944	9.221	.000
Japanese	11.856416	.919174	5294.451	12.899	.000
Korean	11.961041	2.658729	6022.255	4.499	.000
OtherAsian	6.324551	1.890558	4779.886	3.345	.001
Micronesian	-3.366917	1.159567	4932.914	-2.904	.004
Linear	18.713788	.350997	10210.260	53.316	.000
Linear*Disadv	-.570232	.277259	5428.323	-2.057	.040
Linear*Chinese	2.148395	.714155	5208.914	3.008	.003
Linear*Filipino	2.117560	.345426	5186.783	6.130	.000
Linear*OtherAsian	2.966186	1.007106	5109.708	2.945	.003
Linear*Micronesian	1.460766	.649949	5504.644	2.248	.025
Quadratic	-.106814	.094109	8402.541	-1.135	.256

a. Dependent Variable: NWF.

Table 4.9 provides the estimates of the variance components for this model. Once again, the table suggests that in mid-year of Grade K, or initial status for NWF, students were relatively closer together in variance (110.87), but successive measurement occasions indicate they fanned out over time (e.g., 495.6 at time 5 and 469.5 at time 6) as they got closer to the end of first grade. The average correlation within individuals between successive measurements (rho) was 0.294, which indicated a weak-to-moderate average correlation between successive measurements. The between-individual correlation between the intercept and linear time slope [Corr(2,1)] was positive but weak (0.23). This suggests some tendency for students who started higher on NWF in Grade K to make more growth over time than their peers who started lower. The table suggests there was still statistically significant variance in initial status and growth between students that could still be explained ($p < .001$). The baseline variance components in

Model 1 can be used to estimate a reduction in variance in initial status from adding the predictors in Model 2. This can be estimated as follows:

(Model 2 – Model 1)/Model 1

In this case the reduction in variance in initial status would be 251.82 – 205.93= 45.89/251.82 = 0.182 or 18.2%. This suggests the background factors accounted for about 18% of the observed variance in students' initial status score. In contrast, it should be noted the growth variance was modestly larger than the initial model, which can result from slight inaccuracies in estimating the initial variance decomposition model in multilevel modeling (Hox, 2010).

Table 4.9. *Model 1 NWF Variance Components*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Var: [Index1=1]	110.869690	6.615572	16.759	.000
	Var: [Index1=2]	196.448110	5.781960	33.976	.000
	Var: [Index1=3]	234.341260	6.747531	34.730	.000
	Var: [Index1=4]	383.719732	11.452832	33.504	.000
	Var: [Index1=5]	495.632055	16.380465	30.258	.000
	Var: [Index1=6]	469.493159	15.645147	30.009	.000
	ARH1 rho	.293503	.013842	21.204	.000
Intercept + linear [subject = id]	Var(1)	205.930288	8.284371	24.858	.000
	Var(2)	47.267165	2.531681	18.670	.000
	Corr(2,1)	.230587	.039972	5.769	.000

a. Dependent Variable: NWF.

Examining ORF

Model 2, as depicted in Table 4.10, provides the final set of student background variables affecting students' initial status and growth in oral reading fluency. The intercept term indicates the expected value of ORF for the reference group (Caucasian) on its first observation. The table suggests that students who were disadvantaged started at considerably lower levels of ORF (-13.02, $p < .001$) than their more socioeconomically advantaged peers, and female students (7.74,

$p < .001$) started considerably higher than male students. There were a number of initial status differences also associated with race/ethnicity that ranged between -9.58 (Micronesian) and 19.59 (Chinese). In terms of growth, Table 4.8 also suggests disadvantaged students made significantly less growth than their peers (-.338, $p < .001$), and females made significantly more growth than males. Also, growth in ORF initial status associated with race/ethnicity ranged between -1.56 (Chinese) to .922 (Hispanic). Possible interactions between background and the quadratic component in Table 4.8 were also investigated in preliminary models, but one predictor was noted to be statistically significant (Quadratic*Disadv), indicating disadvantaged students fell slightly further behind over time (not tabled). Table 4.8, therefore, indicates the pattern of background influences on student initial status and growth on the two DIBELS measures were very similar.

Table 4.10. *Variance Components for Model 2 ORF*

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	38.118	0.954	6235.673	39.950	0.000
Disadv	-13.019	0.848	5773.123	-15.348	0.000
Female	7.735	0.792	5727.266	9.763	0.000
Chinese	19.597	2.203	5663.652	8.894	0.000
Filipino	7.273	1.216	5961.460	5.983	0.000
Hispanic	-5.791	2.021	6000.639	-2.866	0.004
Japanese	15.368	1.602	5695.625	9.591	0.000
Hawaiian	-4.708	1.097	6017.682	-4.289	0.000
Korean	15.674	3.912	5983.021	4.007	0.000
OtherAsian	14.351	2.967	5880.174	4.837	0.000
Micronesian	-9.583	1.941	6178.700	-4.937	0.000
Samoan	-6.331	2.317	6083.510	-2.733	0.006
Linear	24.361	0.226	10983.236	107.597	0.000
Linear*Disadv	-0.338	0.159	4619.475	-2.132	0.033
Linear*Female	0.357	0.151	4598.466	2.359	0.018
Linear*Chinese	-1.565	0.402	4497.165	-3.891	0.000
Linear*Filipino	-0.684	0.219	4633.324	-3.116	0.002
Linear*Hispanic	0.922	0.385	5069.860	2.393	0.017
Linear*Japanese	-0.708	0.300	4735.990	-2.358	0.018
Linear*Hawaiian	-0.742	0.190	4625.858	-3.901	0.000
Quadratic	-1.824	0.030	6529.769	-61.153	0.000

a. Dependent Variable: ORF

Table 4.11 presents the relevant variance components for Model 2. Relative to Model 1, the initial status variance between students was reduced from 989.54 to 845.74 through the addition of the background predictors. At the initial status for ORF, students were relatively closer together in variance (169.37), but successive measurement occasions indicated they did not disperse much (e.g., 189.82 at time 5, 184.6 at time 6, and 168.26 at time 7) as they got closer to the end of Grade 3. The average correlation within individuals between successive measurements (rho) was 0.32, which suggests a moderate average correlation between successive measurements.

The between-individual correlation between the initial status intercept and linear growth [Corr(2,1)] was negative but weak (-.108). This suggests some tendency for students who started higher in ORF Grade 1 to make less growth over time than their peers who started lower. Once again, the variance components can be used to estimate the reduction in variance from the baseline model (Model 1) to the model with background predictors added. In this case, this was estimated as 0.145 [(989.54 – 845.74) /989.54]. This suggests the background variables accounted for about 14.5% of the variance in students' initial status. The variance in growth was also reduced slightly from 16.11 to 15.91, for a reduction in variance of 1.2% (0.2/16.11, or 0.012). This suggests the background variables accounted for negligible variance in individual growth rates.

Table 4.11. *Variance Components for Model 2 on ORF*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Var: [Index1=1]	169.375595	5.321649	31.828	.000
	Var: [Index1=2]	144.782975	5.471985	26.459	.000
	Var: [Index1=3]	95.844705	3.481606	27.529	.000
	Var: [Index1=4]	183.277290	6.032876	30.380	.000
	Var: [Index1=5]	189.825183	5.887563	32.242	.000
	Var: [Index1=6]	184.602010	5.774088	31.971	.000
	Var: [Index1=7]	168.257821	6.195996	27.156	.000
	Var: [Index1=8]	102.450840	6.750550	15.177	.000
	ARH1 rho	.325154	.014938	21.766	.000
Intercept + time1 [subject = id]	Var(1)	845.736295	18.657122	45.330	.000
	Var(2)	15.905953	.670111	23.736	.000
	Corr(2,1)	-.108013	.021308	-5.069	.000

a. Dependent Variable: ORF.

Model 3: Adding the School Setting in Examining ORF

Model 3 examines how the school settings might also contribute to student learning as measured by ORF. It was determined there was not enough available information about students' schools when they took the NWF measure. It is possible to determine the variance decomposition for three levels regarding ORF (i.e., within students, between students, between schools). The variance components are presented in Table 4.12.

Table 4.12. *Three-level Variance Components for Model 3*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Variance	682.419478	5.652827	120.722	.000
Intercept [subject = schcode]	Variance	215.764533	29.995122	7.193	.000
Intercept [subject = schcode * Rid]	Variance	846.723135	18.015564	47.000	.000

a. Dependent Variable: ORF (mean =67.65, SE = 1.30).

The variance in ORF between schools can be estimated as a ratio of the level-3 variance to the total variance (215.77/1744.91 = 0.124) or 12.4%. The variance between individuals was estimated as 48.5% (846.72/1744.91 = 0.485). The variance within individuals was 39.1%. This provides evidence that a considerable amount of variance in reading scores attributed to individuals on the ORF measure should actually be attributed to differences in the schools the students attend (see Tables 4.4 and 4.6).

Table 4.13 provides the fixed effect estimates of the final school-level model. At this stage, preliminary models examined a range of school-level variables (e.g., school context, enrollment, proportion of teachers who had been at the school for 5 years, average staff teaching experience). There was also one process variable added that in theory would be related to classroom practices (implementation of standards-based curriculum). School-level variables that were not statistically significant ($p < .05$) were dropped from the final model presented in Table 4.13., which shows that only two school-level variables were associated with differences in initial status at the school level. These were the implementation of standards-based curriculum (1.92, $p < .05$) and school context as defined by students receiving free/reduced lunch services, English language services, and special education services (-2.42, $p < .01$).

The first process indicator suggests that a one standard deviation increase in perceived implementation of standards-based curriculum would be related to a 1.92 increase in ORF score. The second suggests a one standard deviation increase in the weighted percentage of students receiving special school services would be associated with a 2.42 decrease in average ORF score. Other background indicators were similar to Model 2. Regarding growth, only school context was significantly associated with differences in school growth in ORF (-0.21, $p < .001$). Once again, there were also associated differences according to race/ethnicity.

Table 4.13. *Fixed Effect Estimates of Background and School Variables Explaining Students' Initial Status and Growth in ORF*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	37.326362	1.180406	647.362	31.622	.000
Schcontext	-2.421638	.839619	148.192	-2.884	.005
Zstandards	1.921036	.807071	152.263	2.380	.019
Disadv	-11.147350	.884478	6081.334	-12.603	.000
Female	8.663685	.776127	6048.874	11.163	.000
Chinese	16.572650	2.318998	6510.660	7.146	.000
Filipino	9.032840	1.272643	6615.739	7.098	.000
Hispanic	-5.465021	2.022023	7222.370	-2.703	.007
Japanese	12.007115	1.703182	6324.835	7.050	.000
Hawaiian	-2.991260	1.162359	6004.431	-2.573	.010
Korean	15.026956	3.994833	6055.522	3.762	.000
OtherAsian	13.124212	3.058182	6015.311	4.292	.000
Micronesian	-7.023554	1.994967	6160.364	-3.521	.000
Linear	24.430896	.177314	28422.237	137.783	.000
Linear*Schcontext	-.211667	.054235	28556.112	-3.903	.000
Linear*Chinese	-1.019020	.258714	28524.057	-3.939	.000
Linear*Filipino	-.792841	.143620	28853.692	-5.520	.000
Linear*Hispanic	1.054466	.257342	28959.078	4.098	.000
Linear*Japanese	-.854838	.193504	28604.310	-4.418	.000
Linear*Hawaiian	-.802858	.124782	28803.732	-6.434	.000
Quadratic	-2.009800	.031477	28285.434	-63.849	.000

a. Dependent Variable: ORF.

Table 4.14 presents the relevant variance components for the school-level model. The reduction in variance can be estimated from a baseline model with just initial status and time in the model. For Level 3, the reduction in variance in initial status was from 108.94 (not tabled) to 40.23 (0.631 or 63.1%). Between individuals, it was from 942.89 (not tabled) to 863.24 (0.084 or 8.4%). For the reduction in growth variance between schools it was 11.02 (not tabled) to 2.90 (0.737, or 73.7%).

Table 4.14. *Model 3 Variance Components*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	Variance	164.352884	1.391076	118.148	.000
Intercept [subject = schcode]	Variance	40.226510	8.447398	4.762	.000
Linear [subject = schcode]	Variance	2.896022	.574709	5.039	.000
Intercept [subject = Rid * schcode]	Variance	863.242462	16.524688	52.240	.000

a. Dependent Variable: ORF.

Model 4: Examining the Relationship Between HSA and ORF

Model 4 examines the relationship between the Hawaii State Assessment (HSA) Grade 3 scores and the ORF Grade 3 scores. In order to make the comparisons, Model 3 was restructured so that initial status (Grade 1) would correspond most closely with the end of Grade 3 (HSA) score. Three models are presented.

The first model (Model 4.1) examines the relationship between the ORF score and the proficiency levels for HSA. This is specified as an ordinal variable with the reference group being category 4 (exceeds proficiency). In Table 4.15, the school ORF intercept (64.23) describes the predicted oral fluency score for students who were in category 4 of HSA (exceeds proficiency). In contrast, the estimates for ORF for HSA category 1 (well below proficiency), category 2 (approaching proficiency) and category 3 (meets proficiency) were all negative, indicating students in those categories had expected ORF scores well below the predicted ORF score (64.23) for students who exceeded proficiency. More specifically, holding other variables constant, the expected ORF score for a student who met proficiency (category 3) was 64.23 – 21.57, or a score of 42.66. A student approaching proficiency (category 2) would have an expected ORF score of 64.23 – 41.57, or 22.66, holding other variables constant. Finally a student well below proficiency (category 1) would have an expected score near zero, holding

other variables constant. This provides one preliminary examination of the correspondence between ORF and Grade 3 reading proficiency scores.

The predictors in Table 4.15 also suggest that school composition (i.e., school context) and stakeholder perceptions of the implementation of standards-based curriculum were not related to Grade 3 ORF scores ($p > .10$), but student composition (i.e., school context) was related to student growth. This suggests a reduction in the growth discrepancies for students attending schools with higher percentages of students receiving free/reduced lunch, English language services, and special education services ($p < .001$).

Table 4.15. *Model 4.1: Examining the Relationship Between HSA Proficiency Scores and ORF, Controlling for School Context and Background Variables*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	64.224990	1.034696	792.167	62.071	.000
Schcontext	-.661588	.711684	151.088	-.930	.354
Zstandards	.853840	.678063	153.451	1.259	.210
[read_pl=1]	-64.320824	1.169519	6183.886	-54.998	.000
[read_pl=2]	-41.573778	.822769	6072.061	-50.529	.000
[read_pl=3]	-21.572223	.767321	5999.431	-28.114	.000
[read_pl=4]	0[b]	0	.	.	.
Disadv	-4.508319	.693136	6043.504	-6.504	.000
Female	2.810278	.607231	6021.890	4.628	.000
Chinese	11.425684	1.862992	7534.026	6.133	.000
Filipino	7.592976	1.024082	7654.518	7.414	.000
Hispanic	-1.763713	1.646213	8608.171	-1.071	.284
Japanese	6.485646	1.367173	7338.664	4.744	.000
Hawaiian	1.174191	.935683	7040.777	1.255	.210
Korean	10.698976	3.089834	6011.446	3.463	.001
OtherAsian	7.582217	2.364062	5950.353	3.207	.001
Micronesian	1.652699	1.558852	6245.991	1.060	.289
Linear	-10.714950	.119386	28973.178	-89.750	.000
Linear*Schcontext	.211313	.043149	28923.727	4.897	.000
Linear*Chinese	.657335	.205318	29086.422	3.202	.001
Linear*Filipino	.589179	.113638	29798.857	5.185	.000
Linear*Hispanic	-.726209	.204130	29977.471	-3.558	.000
Linear*Japanese	.537775	.153539	29307.830	3.503	.000
Linear*Hawaiian	.607688	.098814	29670.670	6.150	.000
Quadratic	-.244787	.029109	28491.126	-8.409	.000

a. Dependent Variable: ORF.

b. This parameter is set to zero because it is redundant.

Model 4.2, as depicted in Table 4.16, examines the relationship between students' Grade 3 HSA scores, expressed as a scaled score (and grand-mean centered), and ORF. In this case, the intercept represents students' final ORF score when the predictors in the model are 0. For the grand-mean centered ORF score (40.55), this score would be interpreted as the expected school

ORF score for schools when the reading scaled score was centered on the sample average for schools. Regarding the HSA scaled score, Table 4.16 suggests a one point increase in reading scaled scores would result in a predicted 0.63 increase in the ORF score, holding other variables in the model constant. This growth model also provided evidence of the predictive relationship between the two measures, HSA scores as outcome and ORF scores as predictors.

Table 4.16. *Model 4.2 Examining the Relationship Between HSA Scaled Scores and ORF*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	T	Sig.
Intercept	40.549008	.945041	572.488	42.907	.000
Schcontext	-.431328	.719387	154.371	-.600	.550
Zstandards	.629082	.683818	158.473	.920	.359
gmread_ss	.661882	.008959	6093.879	73.878	.000
Disadv	-2.932637	.653483	6033.587	-4.488	.000
Female	1.918701	.571746	6015.223	3.356	.001
Chinese	10.312939	1.767626	7834.816	5.834	.000
Filipino	8.594348	.971010	7951.441	8.851	.000
Hispanic	-.975019	1.564373	8921.841	-.623	.533
Japanese	6.583666	1.297013	7660.719	5.076	.000
Hawaiian	2.795218	.889824	7426.783	3.141	.002
Korean	11.888823	2.905078	6004.754	4.092	.000
OtherAsian	7.354886	2.223445	5942.580	3.308	.001
Micronesian	2.887411	1.469613	6273.682	1.965	.049
Linear	-10.700719	.119272	29062.173	-89.717	.000
Linear*Schcontext	.209659	.043113	29026.611	4.863	.000
Linear*Chinese	.652226	.205071	29202.391	3.180	.001
Linear*Filipino	.570474	.113408	29974.523	5.030	.000
Linear*Hispanic	-.715750	.203659	30175.184	-3.514	.000
Linear*Japanese	.517020	.153320	29443.487	3.372	.001
Linear*Hawaiian	.593591	.098629	29834.653	6.018	.000
Quadratic	-.245484	.029096	28532.702	-8.437	.000

a. Dependent Variable: ORF.

Table 4.17, Model 4.3 is a cross-sectional model which examined the association between grand-mean ORF raw scores and HSA scaled scores at a single point in time (Grade 3

ORF EOY 2001 and Grade 3 HSA Spring 2011). In this case, the intercept represents students' final HSA scaled score when the predictors in the model are 0. For the grand-mean centered HSA score (297.71), this score would be interpreted as the predicted school HSA score for schools, where the reading ORF score was at the sample average for participating schools. The average size of the ORF effect on HSA reading scaled score at Grade 3 was approximately 0.68. That means a point increase in oral fluency would be related to a 0.68 point increase in reading scaled scores.

Table 4.17. *Model 4.3: Examining the Relationship Between ORF and HSA Scaled Score*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	Df	t	Sig.
Intercept	297.710440	1.262267	110.538	235.854	.000
schcontext	-3.156662	1.016407	53.567	-3.106	.003
ZEnrollment	-2.700010	.900787	56.162	-2.997	.004
Ndisadv	-5.719718	.878608	3656.019	-6.510	.000
Female	4.025566	.770849	3635.010	5.222	.000
Filipino	-5.130842	1.129011	3648.195	-4.545	.000
Hispanic	-4.921111	1.912404	3625.758	-2.573	.010
Hawaiian	-5.217088	1.009987	3405.777	-5.166	.000
Micronesian	-12.651658	2.010878	3632.912	-6.292	.000
gmoralfluency	.676004	.015719	44.039	43.007	.000

a. Dependent Variable: read_ss.

Table 4.18, Model 4.4, shows that the size of the ORF score varies across schools (Var = 0.005, $p < .07$), which indicates the size of the effect of ORF on reading scaled scores at Grade 3 was varied in size.

Table 4.18. *Model 4.4 Variance Components*

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.
Residual		522.165462	12.380692	42.176	.000
Intercept + gmoralfluency	UN (1,1)	32.192615	10.262193	3.137	.002
[subject = schcode]	UN (2,1)	.012398	.121578	.102	.919
	UN (2,2)	.004670	.002568	1.818	.069

a. Dependent Variable: read_ss.

Table 4.19, Model 4.5, provides an examination of the standardized effect of the HSA reading scaled scores on ORF. Since this is a growth model, ORF represents the outcome variable. The table suggests the standardized effect of the reading HSA score on the ORF score was 0.62 at the school level. This can be interpreted as a one standard deviation increase in the HSA reading scaled score would produce a 0.62 standard deviation increase in the ORF score. This implies a moderate standardized effect.

Table 4.19. *Model 4.5: Examining the Relationship Between the Standardized Grade 3 HSA Score and the ORF Score*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.798743	.016869	137.692	-47.350	.000	-.832098	-.765388
Zread_ss	.620784	.011340	82.644	54.743	.000	.598228	.643340
Linear	-.234722	.000974	28946.978	-240.865	.000	-.236632	-.232812

a. Dependent Variable: Zscore(oralfluency).

Table 4.20, Model 4.6, shows the relationship between the grand mean of the last ORF score (Grade 3 EOY) as the predictor and the grand mean of the Grade 3 HSA reading scaled score as the outcome. Model 4.5 provides an examination of the standardized effect of the ORF on HSA reading scaled scores. The table suggests the standardized effect of the ORF on HSA

reading scaled score was 0.79 at the school level. This suggests a one standard deviation increase in the ORF score would produce a 0.79 standard deviation increase in HSA scaled score. This suggests a strong standardized effect (*r*-square coefficient, 0.62), which implies a relatively strong association between the ORF score and the HSA Grade 3 score.

Table 4.20. *Model 4.6: Examining the Relationship Between the Standardized ORF Score and the Standardized Grade 3 HSA Score*

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	-.564962	.030311	63.885	-18.639	.000
Zoralfluency	.785058	.017254	45.065	45.499	.000

a. Dependent Variable: Zscore(read_ss).

Summary of Major Findings

The findings of this study support the utility of multilevel longitudinal models for examining variations in outcomes within and between individuals and relationships between student background and school variables in contributing to student reading growth during their kindergarten–Grade 3 years. There were several findings of note regarding students' initial reading skills, as measured by NWF and ORF, and the extent to which they acquire reading skills between kindergarten and Grade 3.

First, student scores fanned out relatively quickly as measured on the NWF test between kindergarten and the end of Grade 1. Moreover, the correlation between students' initial status NWF score and growth rate for NWF was positive, suggesting those students who started higher in MOY kindergarten grew more over time than their peers who had lower initial status scores. This information would be useful in targeting academic resources (e.g., extra reading help) early in students' elementary years, before these students fall too far behind their peers.

Second, as students were in school longer, there was more growth in reading that could be attributed to differences between students. This was especially evident by the time students were assessed in terms of ORF. Some of this between-student variation can be attributed to students' socioeconomic background, gender, and race/ethnicity. Evidence of these differences has been referred to as the social distribution of learning (Lee & Bryk, 1989). Model 2 showed similar patterns of background influences on students' initial status and growth as measured by NWF and ORF. On both measures, students who were disadvantaged started at considerably lower levels than their more socioeconomically advantaged peers. Female students also started considerably higher than their male peers. In terms of growth on both measures, disadvantaged students made significantly less growth than their peers, showing an incremental decrease in growth, and females made significantly more growth than males. One key difference in growth between both measures was the acceleration of NWF growth at each occasion for all four ethnic/race subgroups that were kept in the model and the deceleration of ORF growth over time for four of five ethnic/race subgroups that were kept in the model.

Third, students' growth was nonlinear and demonstrated a slight acceleration in growth on the NWF measure and a deceleration of growth on the ORF measure. The results were consistent with previous studies that have found that children's early reading growth is nonlinear (Logan & Petscher, 2010; Nese et al., 2013). The growth curves for the initially high scoring and low scoring students on NWF, however, did not support previous research that suggested that students who initially met thresholds for proficiency on early reading measures, might not be expected to further increase their rate of reading literacy (Ding, 2012). Despite the different growth curves, there was a strong positive relationship between Grade 1 ORF and NWF scores across the first two overlapping test occasions (Grade 1 MOY and Grade 1 end-of-year EOY).

Fourth, the only school-level variables that were significantly associated with differences in initial status at the school level as measured by ORF were implementation of standards-based curriculum and school context (students receiving free/reduced lunch services, language services, and special education services). An increase in perceived implementation of standards-based curriculum was positively related to ORF scores. An increase in the weighted percentage of students receiving special services, as represented by school context, was negatively related to ORF scores. Also, school context was the only school-level variable significantly associated with differences in school growth as measured by ORF. This is not unexpected, since it is assumed that early reading progress is more of a function of family, teacher, and peer influences (Aikens & Barbarin, 2008; Bast & Reitsma, 1998; Burke & Sass, 2008; Hanushek, Kain, Markman, & Rivkin, 2001).

Finally, the last models examined the association between HSA and ORF scores in Grade 3 by employing three different methods, including defining HSA proficiency as an ordinal predictor of reading scaled scores, and also examining the effects of standardized effects associated with the two measures. These results showed a positive association between HSA and ORF measures. More specifically, the standardized effect between the standardized ORF score and the standardized Grade 3 HSA was strong at 0.79. This provides evidence that the oral reading fluency score was a useful indicator of students' HSA scores and can be implemented in Grade 1 as a credible measure through which to monitor student early progress in reading.

Chapter 5 —Discussion, Implications, Conclusions

This chapter presents a discussion and interpretation of the results of the study. The first section summarizes the overview and purpose of the study. The second section discusses the findings in reference to the research questions. The third section draws implications for practice and policy. Finally, the conclusions of the study are presented.

Summary of Study and Purpose

Learning to read is among the most important academic skills that students must acquire during the early years of school (Ding et al., 2013). The growth pattern of children's reading skill development is established early in children's school careers and is known to predict performance in later grades in reading as well as in other areas (Cunningham & Stanovich, 1997; Entwisle, Alexander, & Steffel Olson, 2005; Lonigan et al., 2000). Uncovering the developmental growth trajectories on important fundamental reading outcomes has important academic and clinical implications. A growing number of students enter kindergarten significantly behind their more advantaged and typically developing peers in terms of academic skills, and this performance disparity, particularly between strong and weak readers, becomes increasingly difficult to close (Coyne et al., 2004; Good, Simmons, et al., 2001; Juel, 1988; J. K. Torgesen & Burgess, 1998). Discrepancies in academic growth is perhaps the most important policy issue in education in the United States (Slavin et al., 2009), and although federal legislation aims to "close the achievement gap with accountability" (H. R. 1--107th Congress, 2002, p. 1) as measured by NAEP and high stakes state assessments, there is no national reading test for all students in public schools, specifically that assess children at multiple points in time across kindergarten to Grade 3 on skills identified as foundational in early grades and predictive of future academic success.

Teachers and school leaders need a stronger empirical base to undergird their efforts around screening students on the acquisition of early reading skills and regularly monitoring their development before the results show up in test scores that are administered well after student reading deficits should have been identified and appropriate interventions should have been provided. McCoach et al. (2006) argue that student achievement at specific points in time is a function of prior achievement and growth. Thus, reliable and efficient measures and methods are essential to identify children's early literacy skill development across all primary grades (Fuchs & Fuchs, 1999).

Unlike states that have adopted statewide reading initiatives such as Ohio (Third Grade Reading Guarantee) Arizona (Move on When Reading), Oklahoma (Reading to Succeed), Mississippi (Literacy-Based Promotion Act) and Colorado (Colorado Read Act), Hawaiʻi has not adopted a statewide, cohesive, and comprehensive plan that requires all schools to identify students from kindergarten through Grade 3 who are at-risk for reading difficulties followed by ongoing monitoring to ensure that struggling students receive appropriate instruction designed to address reading deficiencies and accelerate progress. Implementing a statewide plan would allow researchers and practitioners to capture systematic change early and consistently over time across the state, predict future reading performance, investigate reading development from a longitudinal point of view, and explain factors that influence variation and patterns in children's early reading development. The results of this study examining factors that influence children's acquisition of important preliteracy skills at kindergarten entry through Grade 3 provide a deeper exploration of reading growth patterns that can potentially be explained by factors that would not be accounted for without this type of early monitoring of student reading progress.

Purpose

This multilevel, longitudinal, retrospective study was designed to explore growth forms of reading on two DIBELS measures across four-wave longitudinal data, kindergarten through Grade 3 (2008–2011). At the time of this study, there were no other published research identified that employed a multilevel growth modeling approach using the DIBELS NWF and ORF across multiple years relevant to such a relatively large population within a single state. The findings of this study therefore add to the literature about children of diverse demographic backgrounds for whom the available research has remained limited.

This study proposed using growth-modeling methods to examine the relationships between student background and school variables in contributing to students' early literacy growth on two widely used reading measures (NWF and ORF) during their first four years (kindergarten–Grade 3) of primary school. Multilevel growth modeling provides a means to better understand the factors that influence individual variation in development (Muthén, Khoo, Francis, & Boscardin, 2003) and explain variability over time in reading outcomes (Beretvas, 2004) not otherwise detected by means of other statistical methods.

The study examined potential and meaningful differences in children's growth patterns in reading based on their scores as measured by the Dynamic Indicators of Basic Early Literacy Skills 6th Edition (DIBELS) probes for NWF Grades K–2 and ORF Grades 1–3. Next, it examined a set of individual- and school-level factors to determine whether they had any influence on NWF and ORF growth over time. The study utilized a relatively large data set to weigh in on the disparate findings of previous studies aimed at examining the extent to which the developmental dynamics of children's reading performance trajectories follow a cumulative growth model (Bast & Reitsma, 1998; Leppnen et al., 2004; Onatsu-Arvilommi & Nurmi, 2000;

Stanovich, 1986b; Walberg & Tsai, 1983) or a compensatory growth model (Aunola et al., 2004; Jordan et al., 2006).

Conceptual Framework

The conceptual framework underlying this study was primarily based on Creemers and Kyriakides' (2010) dynamic model of educational effectiveness, which proposes a multidimensional approach to measuring school level factors associated with student achievement in various learning outcomes. The view that factors influencing student reading achievement are multilevel is consistent with a broad set of empirical studies directed toward school effectiveness (Tagiuri, 1968; Willms, 1994) as well as research directed more specifically at student literacy (Aaron et al., 2008; Berninger et al., 2004). Against this conceptual backdrop, this study took into account the nested sources of variability between pupils and schools.

Discussion of the Findings

A discussion of general findings of this study is presented in this section, followed by findings specific to the three research questions that guided this study. Broadly speaking, the findings confirm that kindergartners demonstrated a wide disparity in initial reading skills, and these differences or gaps in learning persist during the early years of schooling. This adds to the mounting evidence that children enter kindergarten with differing levels of literacy readiness, and these initial performance levels impact Grade 3 reading performance. Moreover, student- and school-level variables had a differential impact on reading growth. Student-level variables, particularly SES, were better able to explain variability in students' reading development than school-level variables. Based on the reparative perspective, disadvantaged children were not ready at the beginning of school and were more likely to perform at lower levels across school

years. The findings also showed that initial performance predicted rate of growth and that resultant nonlinear growth patterns, whether cumulative or compensatory, tended to differ in terms of the skill being assessed, student background (i.e., gender, race/ethnicity, SES), and students' initial score on a particular measure.

With respect to NWF, students made considerable progress at increasingly varied rates (more the result of individual factors) on this early measure of reading skills during the years, kindergarten–Grade 1, but the rates of growth became more varied over time. Since there was a significant positive relationship between students' initial NWF status and their growth rates, the data suggest those students who started higher in mid-kindergarten increased this advantage over time, compared with their peers who had lower initial status NWF scores. Although this pattern seems to conform to a cumulative model, or Matthew effect, characterized by increasing differences in student performance as well as variance between them (Kempe et al., 2011; Leppnen et al., 2004; Rigney, 2010), there was also considerable difference in the shape of each individual's actual growth trend over time and decreasing predictability of scores across time points.

The proportion of between-student variance in ORF was much larger than that of NWF and growth on the ORF measure slowed down over time. Like NWF, a significantly large proportion of variance in ORF was found to lie between individuals, which supports the proposition that differences in individual students' acquisition of reading skills become more apparent in the earlier years of schooling (Ding, 2012; Foster & Miller, 2007). The slightly negative correlation between the initial status and growth rate indicated a weak tendency for students who had low initial status on ORF to grow slightly more over time and those who had

higher initial status on ORF to make less growth, which was more consistent with a compensatory trajectory (Ding, 2012; Morgan et al., 2011; Salaschek et al., 2014).

Adding school-level variables demonstrated that about 12% of the variance in student scores could be attributed to schools that students attend, but none of the school-level variables appeared to explain initial status and growth rate except the weighted context score and "implementation of standards-based curriculum" at initial status. This suggests that student growth was likely more of a function of classroom variables among others, including family, demographic background, and peer effects.

The findings that follow are specific to this study's three research questions.

1. What is the shape of the average trajectory and the variability from the average as measured by DIBELS NWF and ORF?
2. To what extent is there a relationship between students' scores on (a) DIBELS NWF and ORF measures and (b) Grade 3 DIBELS Oral Reading Fluency test and Hawai‘i State Assessment Grade 3 reading comprehension test?
3. What 2011 (Grade 3) school-level factors additionally contribute to the closing or widening of the reading achievement gap as measured by DIBELS and HSA?

Research Question 1

The first research question examined the patterns of performance that distinguish various subgroups of students. Thus, this study also examined whether initial differences remain stable, converge, or diverge. In terms of NWF, this study found that the correlation between students' initial status NWF score and growth rate for NWF was positive, suggesting students who had higher initial status NWF scores in kindergarten demonstrated more growth over time than their peers who started with lower initial status NWF scores in kindergarten. Some of this between-

student variation can be attributed to student characteristics (i.e., socioeconomic background, gender, and race/ethnicity) as demonstrated by their influences on students' initial status and growth measured by NWF and ORF. On both measures, the initial status of the disadvantaged (or low SES) subgroup and male subgroup had considerably lower levels than their counterparts. In terms of growth on NWF and ORF, the disadvantaged subgroup and male subgroup made significantly less growth than their counterparts. One key difference in growth between both measures, however, was the acceleration of growth on NWF at each occasion for all four ethnic/race subgroups (Chinese, Filipino, Other Asian, Micronesian) that were kept in the model and the deceleration of growth on ORF over time for four of five ethnic/race subgroups (Chinese, Filipino, Japanese, and Hawaiian) that were kept in the model, compared with the Caucasian reference group.

Research Question 2

The second research question examined the extent of the relationship between students' scores on (a) DIBELS NWF and ORF measures and (b) Grade 3 DIBELS Oral Reading Fluency test and Hawai‘i State Assessment Grade 3 reading comprehension test. After establishing the pattern of growth and variability in children's reading development on NWF (kindergarten–Grade 2) and ORF (Grade 1–Grade 3), this study analyzed the concurrent correlations between raw scores of both measures across the first two overlapping test occasions (i.e., Grade 1 Middle of the Year and End of the Year scores). Based on most standards, there was a strong positive correlation observed for the first two test occasions shared between the NWF and ORF measures, suggesting a good validity correspondence in monitoring the acquisition of early reading skills over time.

Different methods were employed to examine the association between HSA and ORF scores in Grade 3, each of which demonstrated a positive association between HSA and ORF measures. The strong standardized effect between the standardized ORF score and the standardized Grade 3 HSA provided strong supporting evidence that the ORF score was a useful predictor of students' HSA scores and can, therefore, be considered as a credible measure through which to monitor students' early progress in reading.

Research Question 3

The third research question investigated school-level factors that potentially contribute to the closing or widening of the reading achievement gap as measured by DIBELS and HSA. The study found just a few school-level variables that were significantly associated with differences in initial status at the school level as measured by ORF. Among the variables examined were school personnel perceived (a) implementation of standards-based curriculum and (b) provision of student support services or school context (i.e., differentiated classroom practices and personalized classroom climate; prevention/early intervention; family participation; support for transition; community outreach and support; specialized assistance; and crisis emergency support).

An increase in perceived implementation of standards-based curriculum was positively related to school-level ORF scores, while an increase in the weighted percentage of students receiving special services, as represented by school context, was negatively related to ORF scores. Additionally, school context was the only school-level variable significantly associated with differences in school growth as measured by ORF. This was expected, since reading progress is assumed to be more of a function of influences related to family, teacher, and peer

than a sole function of school-level variables (Adams, 1990; Mastropieri et al., 1999; McNamara et al., 2011).

Limitations of the Study

Although multilevel growth modeling shows promise as an effective means for examining the relationships between student background and school variables in contributing to students' early literacy growth, there are a number of limitations relevant to this study to keep in mind in interpreting the results. First, although this study included a large student pool, the students represented only elementary schools in Hawai‘i that used DIBELS, which might limit its generalizability to the total population of elementary schools in the state and elsewhere. Second, because classroom variables were not included in this study, variance due to classroom factors may be attributed to students or to schools. Therefore, while it may appear that nearly all of the variance (for example, 60% between student variance for ORF) is attributable to between-student differences, this statistic also includes any variance that could be due to between-classroom differences. Hence, the amount of variance due to classrooms within schools is not known. A third possible limitation of the study concerns the level of integrity at which assessors (e.g., classroom teachers and other school staff) administered the DIBELS assessments. This possible limitation, however, poses the least impact to the study given the objective and straight forward nature of the probes. Finally, since Good and Kaminski (Good & Kaminski, 2002a) determined that the readability formula explains only about a third of the variance in student performance on ORF, there are likely other factors that influence students' performance across grades. It remains for further research to refine both the readability as well as locate other factors that affect performance on ORF.

Implications

Despite the outlined limitations, the study has several implications for policy, practice, and further research. The implications of this study are important to consider, given that efforts to decrease the overall achievement gap in reading development, locally and nationally, is still a pervasive concern (National Center for Education Statistics, 2013a; Pfost et al., 2014; Shin et al., 2013). Although the findings of this study were based on extant data from the DIBLES 6th Edition (i.e., a previous iteration that was revised in 2011) and a previous version of Hawai‘i’s state assessment reading subtest, they shed light on the utility of the DIBELS early NWF and ORF subtests to monitor students’ reading development over time using a multilevel and longitudinal analytic approach on a large student and school data set.

Implications for Practice and Policy

Consistent with previous research (Boscardin et al., 2008; Francis et al., 1996) and the results of this study, reading difficulties are characterized by deficits in foundational reading skills that can be identified as early as kindergarten. This study accentuates the need for early identification of reading difficulties coupled with interventions, initiated at the earliest point in time, specifically targeting deficits in prerequisite reading skills. DIBELS NWF initial status data show that there are large discrepancies in reading skills when children enter kindergarten, and subsequent NWF and ORF scores show that gaps persist as students move to subsequent grades.

Given the strong positive correlation observed between NWF and ORF for the first two common test occasions, schools should not neglect teaching children skills relevant to these measures, which are necessary to acquiring automaticity of decoding skills across early grades and important determinants of reading success (National Reading Panel, 2000). Additionally, the Spring 2011 Grade 3 DIBELS ORF results correlated strongly with, and provided significant

predictive ability toward, reading comprehension achievement as measured on the 2011 state-mandated Grade 3 reading subtest. This provides support for the continued practice of using DIBELS ORF assessments to monitor early student progress in reading across Grades 1–3, as the data provides predictive power toward reading comprehension achievement. More broadly, DIBELS, as an interim assessment, should be considered to be a part of HIDOE's comprehensive assessment system used in tandem with formative and summative assessments. Furthermore, although skills associated with DIBELS measures are foundational to future reading success, meeting minimum competency on these skills should not be equated with reading proficiency.

The implication of this study in terms of the contribution of school setting is not as strong, but warrants attention. Since an increase in perceived implementation of standards-based curriculum was positively related to initial ORF scores, it makes a case for the benefits of increasing teachers' understanding of and effectively addressing standards in the classroom as a potentially effective practice for improving reading outcomes. Another school-level implication might be to ensure that sufficient time and resources are being invested in targeted instruction that addresses skills associated with ORF given that an increase in the weighted percentage of students receiving special services. This is because the results indicated that student composition variables were negatively related to school ORF scores. Student composition is likely a proxy indicator for other types of educational processes that may differ across school contexts (e.g., access to quality teaching resources, academic support services, community and parent expectations). Although the contribution of school setting was not strong, the flattening of the growth slope as measured by NWF and ORF following the spring end-of-year DIBELS test occasions each year (as identified in preliminary growth models) implicates the efficacy of school and draws attention to the fact that summer break is a possible source of achievement

shortfalls experienced in primary grades by certain subgroups of students, particularly those of lower SES backgrounds in particular types of school contexts.

At the policy level, a strong educational plan in reading includes reciprocity between curriculum, assessment, and instruction (Ransom, Santa, Williams, & Farstrup, 1999). While the reciprocity of curriculum, assessment, and instruction is critical, exactly what defines these components will be guided by policy initiatives such as the Common Core State Standards (CCSS) released in 2010 by the National Governors Association (NGA) Center for Best Practices and the Council of Chief State School Officers (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Although these new internationally benchmarked and college and career focused standards are being aligned to formative instructional practice that will include the integration of literacy components (reading, writing, listening, speaking, and language), DIBELS assessments, including its most recent Edition (DIBELS Next), assesses basic early literacy skills that are delineated in the CCSS for English Language Arts (ELA). Ensuring progress of our most academically and socioeconomically disadvantaged children will require continued monitoring of these important and meaningful reading outcomes (National Reading Panel, 2000).

This study supports the utilization of rigorous formative assessment measures as recommended in NCLB (H. R. 1--107th Congress, 2002) and the proposed reauthorization of ESEA (U.S. Department of Education, 2010). It extends the evidence for using student performance on measures such as NWF and ORF as a means for instructional decision making. Systematic monitoring of children's early reading skill acquisition using reliable measures is critical for identifying student needs and informing instruction. DIBELS was designed to facilitate and guide instructional support tailored to student needs. According to the authors of

DIBELS, the measures serve as a decision-making tool within a framework that determines children's response to instruction and therefore, should not be used in isolation or as an instrument for high-stakes decision making (Dynamic Measurement Group, 2007).

This study also lends support for examining children's reading growth over time, longitudinally. Student achievement at any given point is a function of both prior student achievement and student growth (McCoach et al., 2006). Schleicher (2011) argued that a shortcoming of NCLB legislation was its annual "single-bar" assessment, which focused unduly on meeting proficiency and thereby not valuing achievement growth. He noted that assessment and accountability systems used by high performing countries incorporate progressive learning targets that delineate learner's level of proficiency relevant to incremental levels of progress on a pathway across grade levels (Schleicher, 2011). A snapshot of student achievement at a single time point is likely to depict a different narrative of school success and effectiveness compared to a model that captures students' growth over time (McCoach et al., 2006). Relative to this study, a single measure would not have revealed differential growth patterns of student subgroups and the extent to which the reading achievement gaps might differ among the subgroups across time. Although effects of possible interventions cannot be inferred by this study, the findings imply, in accordance with research, that collecting data for multilevel analyses and investigation of reading trajectories should be coupled with ongoing examination of the data in reference to benchmark goals and the implementation of instructional strategies designed to remediate or prevent reading difficulties.

Implications for Future Research

In light of the research questions, a multilevel model was appropriate for this study, as it permitted the examination of multiple levels of data concurrently to explain variability in early

reading outcomes. Research is still needed that can examine other processes that influence children's early reading outcomes within schools. One set of variables might be classroom contexts within the schools (e.g., teacher skills, reading instruction, and monitoring) which were variables that were outside the data available for use in this study.

In terms of school context variables, other school resources and processes (e.g., core reading program, reading incentives, and summer reading programs) should be further examined to determine the extent of their influence on the outcomes. Also, given the disparity in the initial kindergarten NWF scores, it can be assumed that children's pre-kindergarten literacy experiences vary in terms of the extent to which children learn pre-literacy skills, and therefore an extension of this study can incorporate either student- or classroom-level data that are amenable to multilevel analysis. For example, since 2002, HIDOE kindergarten teachers have administered the Hawaiʻi State School Readiness Assessment (HSSRA), which is a classroom-level assessment that elicits teachers' perceptions, based on beginning-of-year observations of their kindergarten students on five developmental domains, one of which is Academic-Literacy Concepts and Skills (Hawaiʻi Department of Education, Systems Accountability Office, System Evaluation & Reporting Section, 2013). This particular domain comprises four descriptive items (knows names and sounds of more than three letters; uses symbols, scribbles or letter-like form to "write words"; shows familiarity with how books work; communicates ideas and describes things using phrases and sentences) on which the teacher rates on a five-point Likert scale. The elementary school principals also complete another version of the HSSRA that gauges principals' perceptions of their respective school's policies and practices that support successful early learning.

In conjunction with the U.S. Department of Health and Human Services, the U.S. Department of Education is supporting President Obama's administration's proposal for new investments to support a continuum of high-quality early learning for pre-kindergarten children. This initiative will prompt states to begin investing more resources in measuring early learning impact. Including pre-kindergarten measures that potentially contribute to later reading acquisition might further explain reading trajectories depicted in this study.

This study's preliminary findings showed that slopes flattened after summer indicating that summer setback, or seasonal patterns of learning, may play a critical role in the reading achievement gap. It would be interesting to investigate the extent to which reading-achievement gaps disperse as a result of differential growth rates during summer. Also, a closer examination of the extent to which summer setback contributes to subsequent growth patterns of different subgroups of children would enable educational leaders and policy makers to consider calendar reforms or the provision of classroom- and home-based summer enrichment experiences, especially for those identified as at-risk for reading difficulties.

Using measures amenable to longitudinal examination serves to reconceptualize early reading assessment practices for identifying children with reading difficulties and opens new research possibilities in this area. Future research should continue to model the multiple levels of school conditions that influence children's reading development over substantial periods of time.

Conclusions

The results of the study add to the incremental process of knowledge building in early student literacy. The findings support the potential utility of DIBELS NWF and ORF as a means to explicate and map students' reading growth trajectories across multiple years. Additionally, the study adds to the small but growing body of research that apply a multilevel frame to

investigate the relationship between student background and school variables in contributing to students' reading development and patterns of growth. There is great potential for earlier assessment and intervention in a more systematic way as a means of investigating factors that contribute to students' early literacy development beyond a single point in time. Other researchers are encouraged to explore the potential of these varied models coupled with intervention research as a means of clarifying and expanding our understanding of the relationship between early intervention in reading and student learning over time. This challenge remains relevant and urgent as both researchers and practitioners seek best practices and tools to measure and improve reading performance across time. This line of inquiry in the future could result in more-grained information for policy and practice concerning the formulation of contextualized early reading resources, interventions, and monitoring, for targeted subgroups of students.

References

Aarnouse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation: An International Journal on Theory and Practice, 6*(3), 251-278.

Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the componenet model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities, 41*(1), 67-84.

Abbott, M., Wills, H., Miller, A., & Kaufman, J. (2012). The relationship of error rate and comprehension in second and third grade oral reading fluency. *Reading Psychology, 33*, 104-132.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*(2), 235-251. doi: 10.1037/0022-0663.100.2.235

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171-191.

Allington, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Allington, R. L., & McGill-Franzen, A. (2003). The impact of summer setback on the reading achievement gap. *The Phi Delta Kappan, 85*(1), 68-75.

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., . . . Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology, 31*(5), 411-427.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*(5), 255-259.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. *Journal of Educational Psychology, 96*(4), 699-713.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame‘enui, E. J., & Beck, C. T. (2008). Reading Fluency as a Predictor of Reading Proficiency in Low-Performing, High-Poverty Schools. *School Psychology Review, 37*(1), 18-37.

Barger, J. (2003). Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment Ashville, NC: North Carolina Teacher Academy.

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago, IL: University Chicago Press.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*(6), 1373-1399. doi: 10.1037/0012-1649.34.6.1373

Batsche, G. M., Kavale, K. A., & Kovaleski, J. F. (2006). Competing views: A dialogue on Response to Intervention. *Assessment for Effective Intervention, 32*(1), 6-19.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506-521.

Beech, J. R. (2005). Ehri's model of phases of learning to read: a brief critique. *Journal of Research in Reading, 28*(1), 50-58. doi: 10.1111/j.1467-9817.2005.00252.x

Below, J. L., Skinner, C. H., Fearrington, J. Y., & Sorrell, C. A. (2010). Gender differences in early literacy: Analysis of kindergarten through fifth-grade Dynamic Indicators of Basic Early Literacy Skills Probes. *School Psychology Review, 39*(2), 240-257.

Beretvas, S. N. (2004). Using hierarchical linear modeling for literacy research under No Child Left Behind. *Reading Research Quarterly, 39*(1), 95-99.

Berninger, V. W., Dunn, A., Lin, S. J. C., & Shimada, S. (2004). School evolution scientist—practitioner educators creating optimal learning environments for all students. *Journal of Learning Disabilities, 37*(6), 500-508.

Bethell, C., Forrest, C., Stumbo, S., Gombojav, N., Carle, A., & Irwin, C. (2012). Factors promoting or potentially impeding school success: Disparities and state variations for children with special health care needs. *Maternal & Child Health Journal, 16,* 35-43. doi: 10.1007/s10995-012-0993-z

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31*(7), 1027-1050.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. . *Journal of Research on Educational Effectiveness, 1,* 289-328.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal, 108*(2), 115-130.

Boland, T. (1993). The importance of being literate: Reading development in primary school and its consequences for the school career in secondary education. *European Journal of Psychology of Education, 8*(3), 289-305.

Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis, 18*, 309-326.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*(1), 192-208.

Bowey, J. A. (1994). Syntactic awareness in relation to reading skills and on-going reading comprehension monitoring. *Journal of Experimental Psychology, 41*, 282-299.

Brown, A. L., Palincsar, A. S., & Purcell, L. (1986). Poor readers: Teach, don't label. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (pp. 105-143). Hillsdale, NJ: Erlbaum.

Buck, J., & Torgesen, J. K. (2003). The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test Tallahassee, FL: Florida Center for Reading Research.

Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis, 24*(3), 219-239.

Burke, M. A., & Sass, T. R. (2008). Classroom peer effects and student achievement: Federal Reserve Bank of Boston.

Butler, S. R., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1983). A seven year longitudinal study of early prediction of reading achievement. *Journal of Educational Psychology, 77*, 349-361.

Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology, 102*(2), 257-280.

Carlson, E., Jenkins, F., Li, T., & Brownell, M. (2013). The interactions of vocabulary, phonemic awareness, decoding, and reading comprehension. *Journal of Educational Research, 106*(2), 120-131. doi: 10.1080/00220671.2012.687791

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*, 331-361.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects sssociated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163-176.

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163-176. doi: 10.1177/0022219408326219

Chall, J. S. (1983). *Stages of reading development* (2 ed.). Forth Worth, TX: Harcourt-Brace.

Chaney, C. (1992). Language development, metalinguistic skills, and print awareness in 3-year-old children. *Applied psycholinguistics, 13*, 485-514.

Chard, D. J. (1995). Understanding the primary role of word recognition in the reading process: Synthesis of research on beginning reading. Technical report No. 15. National Center to Improve the Tools of Educators, College of Education: University of Oregon.

Chard, D. J., Stoolmiller, M., Harn, B. A., Wanzek, J., Vaughn, S., Linan-Thompson, S., & Kame‘enui, E. J. (2008). Predicting reading success in a multilevel schoolwide reading model a retrospective analysis. *Journal of Learning Disabilities, 41*(2), 174-188.

Children’s Educational Services. (1987). Test of Reading Fluency. Minneapolis, MN: Author.

Clemens, N. H., Hilt-Panahon, A., Shapiro, E. S., & Yoon, M. (2012). Tracing student responsiveness to intervention with early literacy skills indicators: Do they reflect growth toward text reading outcomes? *Reading Psychology, 33*(1/2), 47-77. doi: 10.1080/02702711.2011.630608

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly, 26*(3), 231-244. doi: 10.1037/a0025173

Coyne, M. D., & Harn, B. A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools, 43*(1), 33-43.

Coyne, M. D., Kame‘enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten Intervention. *Journal of Learning Disabilities, 37*(2), 90-104.

Creemers, B., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research, 54*(3), 263-294.

CTB/McGraw-Hill. (2004). TerraNova, The (2nd ed.) (CAT/6). Monterey, CA: CTB/McGraw-Hill.

Cummings, K. D., Dewey, E. N., Latimer, R. J., & Good, R. H. (2011). Pathways to word reading and decoding: The roles of automaticity and accuracy. *School Psychology Review, 40*(2), 284-295.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934-945.

Demont, E., & Gombert, J. E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology, 66*(3), 315-332.

Deno, S., Fuchs, L. S., Marston, D., & Shinn, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-524.

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45*(3), 232-243. doi: 10.1177/0022219412442155

Denton, K., & West, J. (2002). Children's reading and mathematics achievement in kindergarten and first grade Washington, DC: National Center for Education Statistics.

Dickinson, D., & Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school* Baltimore, MD: Brookes.

Ding, C. (2012). Studying children's early literacy development: Confirmatory multidimensional scaling growth modeling. *International journal of Educational Research, 53*, 278-288.

Ding, C., Richardson, L., & Schnell, T. (2013). A developmental perspective on word literacy from kindergarten through the second Grade. *Journal of Educational Research, 106*(2), 132-145.

Downey, D. B., von Hippel, P. T., & Beckett, A. B. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*, 613-635.

Duncan, G. J., Classens, A., Huston, A. C., Pagani, L. S., Engel, M., Sexton, H., . . . Duckworth, K. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.

Dynamic Measurement Group. (2007). DIBELS: Myths and facts. Eugene, OR: Author.

Echols, L. D., West, R. F., Stanovich, K. E., & Zehr, K. S. (1996). Using children's literacy activities to predict growth in verbal cognitive skills: A longitudinal investigation. *Journal of Educational Psychology, 88*, 296-304.

Ehri, L. C. (1995). Phases of development in learning to read by sight. *Journal of Research in Reading, 18*, 116-125.

Ehri, L. C., & Wilce, L. S. (1983). Development of word identification speed in skilled and less skilled beginning readers. *Journal of Educational Psychology, 75*, 3-18.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20*, 163-179.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition and math achievement in the first two years of school. *American Sociological Review, 57*, 72-84.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Boulder, CO: Westview Press.

Entwisle, D. R., Alexander, K. L., & Steffel Olson, L. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology, 110*(5), 1458-1502.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234-248.

Field, A. (2009). *Discovering statistics using SPSS*. Los Angeles, CA: SAGE.

Fien, H., Yonghan, P., Baker, S., Mercier Smith, J. L., Stoolmiller, M., & Kame‘enui, E. J. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review, 39*(4), 631-653.

Fiester, L. (2013). Early warning confirmed: A research update on third-grade reading. Baltimore, MD: The Annie Casey Foundation.

Foster, W. A., & Miller, M. (2007). Development of the literacy achievement gap: A longitudinal study of kindergarten through third grade. *Language, Speech, and Hearing Services in Schools, 38*(3), 9.

Francis, D. J., & Fletcher, J. M. (1996). Defining learning and language disabilities: Conceptual and psychometric issues with the use of IQ tests. *Language, Speech & Hearing Services in Schools, 27*(2), 132-143.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students’ oral reading fluency using DIBELS. *Journal of School Psychology, 46*(3), 315-342. doi: 10.1016/j.jsp.2007.06.003

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3-17.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301-330). Hillsdale, NJ: Erlbaum.

Fuchs, L. S., & Fuchs, D. (1999). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7-21.

Gillon, G., & Dodd, B. J. (1994). A prospective study of the relationship between phonological, semantic and syntactic skills and specific reading disability. *Reading and Writing: An Interdisciplinary Journal, 6*(4), 321-345.

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3), 463-483.

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. *Best Practices in School Psychology, IV*(1), 699-720.

Good, R. H., & Kaminski, R. A. (2002a). DIBELS Oral Reading Fluency passages for first through third grades (Technical Report No. 10). Eugene, OR: University of Oregon.

Good, R. H., & Kaminski, R. A. (2002b). Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Retrieved October 27, 2014, from http://dibels.uoregon.edu

Good, R. H., & Kaminski, R. A. (2003). DIBELS: Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Longmont, CO: Sopris West.

Good, R. H., & Kaminski, R. A. (n.d.). What are DIBELS? Eugene, OR: Dynamic Measurement Group.

Good, R. H., Kaminski, R. A., Smith, S., & Bratten, J. (2001). Technical adequacy of second grade DIBELS oral reading fluency passages (Technical Report, No. 8). Eugene, OR: University of Oregon.

Good, R. H., Simmons, D. C., & Kame‘enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(257-288).

Good, R. H., Wallin, J., Simmons, D. C., Kame‘enui, E. J., & Kaminski, R. A. (2002). System-wide percentile ranks for DIBELS benchmark assessment *Technical Report 9*. Eugene, OR: University of Oregon.

Goodman, K. S. (2006a). A critical review of DIBELS. In K. Goodman (Ed.), *Examining DIBELS: what it is and what it does*. Brandon, VT: Vermont Society for the Study of Education.

Goodman, K. S. (2006b). The truth about DIBELS. Portsmouth, NH: Heinemann.

Gough, P. B., & Juel, C. (1991). The first stages of word recognition. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 47-56). Hillsdale, NJ: Erlbaum.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Gunning, T. G. (1998). *Assessing and correcting reading and writing difficulties*. Boston: Allyn & Bacon.

H. R. 1--107th Congress. (2002). *No Child Left Behind Act of 2001*. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Hall, S. L. (2012). *I've DIBEL'd, now what? Next Edition*. Longmont, CO: Cambium Learning Group.

Handler, S. M., & Fierson, W. M. (2011). Learning disabilities, dyslexia, and vision. *Pediatrics, 127*(3), 818-856. doi: 10.1542/peds.2010-3670

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2001). Does peer ability affect student achievement?

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008a). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41*(2), 143-157.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008b). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41*, 143-157. doi: 10.1177/0022219407313585

Hawai‘i Department of Education. (2013). Official enrollment count 2013–2013 [Data file]. Retrieved October 25, 2014, from http://www.hawaiipublicschools.org/ConnectWithUs/MediaRoom/PressReleases/Pages/Hawaii-public-school-enrollment-grows-by-2%2c000-students.aspx

Hawai‘i Department of Education. (2014). Strive HI performance system: SY 2013-2014 statewide snapshot. Retrieved September, 2014, from http://www.hawaiipublicschools.org/DOE%20Forms/StriveHI2014/state_snapshot_2014.pdf

Hawai‘i Department of Education, Systems Accountability Office, System Evaluation & Reporting Section. (2013). Hawaii State School Readiness Assessment. Retrieved October 23, 2014, from http://arch.k12.hi.us/school/hssra/hssra.html

Hayes, D. P., & Grether, J. (1983). The school year and vacations: When do students learn? *Cornell Journal of Social Relations, 17*(1), 56-71.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Mulitlevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development, 58*, 1151-1160.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172-177.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Pschology in the Schools, 43*, 45-56.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnositc accuracy of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*(4), 541-556.

Hock, M. F., Brasseur, I. F., Deshler, D. D., Catts, H. W., Marquis, J. G., Mark, C. A., & Stribling, J. W. (2009). What Is the reading component skill profile of adolescent struggling readers in urban schools? *Learning Disability Quarterly, 32*(1), 21-38.

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 14*, 265-274. doi: 10.1016/s0022-5371(75)80070-3

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40*(1), 108-131.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Invernizzi, M., Justice, L., Landrum, T., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36*(4), 479-500.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure: Toward an integrated theory of atypical reading*. Philadelphia, PA: Psychology Press.

Johnson, R. S. (2002). *Using data to close the achievement gap: How to measure equity in our schools*. Thousand Oaks, CA: Corwin Press.

Johnston, R., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal, 8*, 217-234.

Joint Commettee of the Hawaii State Teachers Association and State of Hawaiʻi Department of Education. (2014). Joint committee recommendations regarding educator effectiveness system design and implementation. Honolulu, HI: Hawaiʻi Department of Education, Educator Effectiveness System.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153-175. doi: 10.1111/j.1467-8624.2006.00862.x

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Kameʻenui, E. J., & Carnine, D. W. (1998). *Effective teaching strategies that accommodate diverse learners*. Des Moines, IA: Prentice-Hall Inc.

Kame‘enui, E. J., Francis, D. J., Good, R. H., O’Connor, R. E., & Simmons, D. C. (2002). An analysis of reading assessment instruments for K–3. University of Oregon: Institute for the Development of Educational Achievement.

Kame‘enui, E. J., Fuchs, L., Francis, D. J., Good, R., III, O’Connor, R. E., Simmons, D. C., . . . Torgesen, J. K. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35*(4), 3-11.

Kaminski, R. A., Good, R. H., Baker, D., Cummings, K., Dufour-Martel, C., Peterson, K., . . . Wallin, J. (2008). Position paper on use of DIBELS for student-level accountability decisions. Eugene, OR: Dynamic Measurement Group.

Kellas, G., Ferraro, R. F., & Simpson, G. B. (1998). Lexical ambiguity and the time-course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 601-609.

Kempe, C., Eriksson-Gustavsson, A.-L., & Samuelsson, S. (2011). Are there any Matthew Effects in literacy and cognitive development? *Scandinavian Journal of Educational Research, 55*(2), 181-196. doi: 10.1080/00313831.2011.554699

Kent, S. C., Wanzek, J., & Al Otaiba, S. (2012). Print reading in general education kindergarten classrooms: What does it look like for students at-risk for reading difficulties? *Learning Disabilities Research & Practice (Wiley-Blackwell), 27*(2), 56-65. doi: 10.1111/j.1540-5826.2012.00351.x

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in Oral Reading Fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*(3), 652-667.

Klecan-Aker, J. S., & Caraway, T. H. (1997). A study of the relationship of storytelling ability and reading comprehension in fourth and sixth grade African-American children. *International Journal of Language & Communication Disorders, 32*(1), 109-125.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*, 172-192.

Lentz, F. E. J. (1988). On-task behavior, academic performance, and classroom disruptions: Untangling the target selection problem in classroom interventions. *School Psychology Review, 17*, 243-257.

Leppnen, U., Niemi, P., Aunola, K., & Nurmi, J.-E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly, 39*(1), 72-93.

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004). Reading performance and its developmental trajectories during the first and the second grade. *Learning and Instruction, 14*(2), 111-130. doi: http://dx.doi.org/10.1016/j.learninstruc.2004.01.006

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47*(3), 596-632.

Levine, L. (1977). Differentiation of letter-like forms in prereading children. *Developmental Psychology, 13*, 89-94.

Logan, J. A. R., & Petscher, Y. (2010). School profiles of at-risk student concentration: Differential growth in oral reading fluency. *Journal of School Psychology, 48*(2), 163-186. doi: 10.1016/j.jsp.2009.12.002

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596-613.

Lonigan, C. J., Schatschneider, C., & Westberg, L. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling *Developing early literacy: A scientific synthesis of early literacy development and implications for intervention*. Jessup, MD: National Institute for Literacy & The Partnership for Reading.

Luft Baker, D., Park, Y., & Baker, S. K. (2010). Effect of initial status and growth in pseudoword reading on Spanish reading comprehension at the end of first grade. *Psicothema, 22*(4), 955-962.

Lyon, G. R. (1998). Overview of reading and literacy initiatives. Washington, DC: National Institute of Child Health and Human Development.

Lyon, G. R. (2002). Reading Development, Reading Difficulties, and Reading Instruction Educational and Public Health Issues. *Journal of School Psychology, 40*(1), 3.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., & Wood, F. B. (2001). Rethinking learning disabilities. In J. C. E. Finn, R. A. J. Rotherham & J. C. R. Hokanson (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3, pp. 199-221). New York, NY: Academic Press.

Mason, J. M. (1980). When do children begin to read: An exploration of four-year-old children's letter and word reading competencies. *Reading Research Quarterly, 15*, 203-227.

Mastropieri, M. A., Leinart, A., & Scrugges, T. E. (1999). Strategies to increase reading fluency. *Intervention in School and Clinic, 35*, 278-283, 292.

McCoach, D. B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in psychology, 1*(17), 1-12.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*(1), 14-28. doi: 10.1037/0022-0663.98.1.14

McNamara, J. K., Scissons, M., & Gutknecth, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of Learning Disabilities, 44*(5), 421-430.

Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly, 31*(4), 187-198.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*(5), 417-436.

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472-488.

Morris, D., Bloodgood, J. W., Lomax, R. G., & Perney, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly, 38*(3), 302-328.

Muthén, B., Khoo, S., Francis, D. J., & Boscardin, C. K. (2003). Analysis of reading skills development from kindergarten through first grade: An application of growth mixture

modeling to sequential processes. In N. Duan & S. Reise (Eds.), *Multilevel modeling: Methodological issues and applications* (pp. 71-89). Mahwah, NJ: Erlbaum.

National Center for Education Statistics. (2013a). A first look: 2013 mathematics and reading. National assessment of educational progress at grade 4 and 8. Washington, DC: National Center for Education Statistics.

National Center for Education Statistics. (2013b). The Nation's report card: Reading 2013 state snapshot reports for grade 4. Retrieved April 14, from U.S. Department of Education

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects. Retrieved April 4, 2014, from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

National Institute for Literacy. (2008). Developing early literacy: Report of the National Early Literacy Panel. A scientific synthesis of early literacy development and implications for intervention. 260.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups: National Institute of Child Health and Human Development, National Institutes of Health.

Nese, J. F. T., Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across Grades 1–8. *Journal of School Psychology, 51*(5), 625-642. doi: 10.1016/j.jsp.2013.05.006

Nese, J. F. T., Park, B. J., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *The Elementary School Journal, 111*(4), 608-624.

Neuman, S. B., & Dickinson, D. K. (2001). *Handbook of early literacy research*. New York, NY: Guilford.

Nicholson, T. (1997). Closing the gap on reading failure: Social background, phonemic awareness, and learning to read. *Foundations of Reading Acquisition and Dyslexia: Implications for Early Intervention*, 381-407.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disability in kindergarten and first grade. *Scientific Studies of Reading, 3*, 159-197.

Oka, E., & Paris, S. (1986). *Patterns of motivation and reading skills in underachieving children* (Vol. 2). Hillsdale, NJ: Erlbaum.

Onatsu-Arvilommi, T., & Nurmi, J.-E. (2000). The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A Cross-lagged longitudinal study. *Journal of Educational Psychology, 92*(3), 478-491.

Ortiz, M., Folsom, J. S., Otaiba, S. A., Greulich, L., Thomas-Tate, S., & Connor, C. M. (2012). The componential model of reading: Predicting first grade reading performance of culturally diverse students from ecological, psychological, and cognitive factors assessed at kindergarten entry. *Journal of Learning Disabilities, 45*(5), 406-417.

Paleologos, T. M., & Brabham, E. G. (2011). The effectiveness of DIBELS oral reading fluency for predicting reading comprehension of high- and low-income students. *Reading Psychology, 32*(1), 54-74.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184-202.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, *97*(3), 299-319.

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357-383.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, *28*(3), 5-13.

Peugh, J. L., & Enders, C. K. (2004). Missing data in Educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525-556.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew Effects in reading. *Review of Educational Research*, *84*(2), 203-244. doi: 10.3102/0034654313509492

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology*, *94*(1), 3-13.

Powell, D., Plaut, D., & Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child? *Journal of Research in Reading*, *29*(2), 229-250.

Pressley, M., Hilden, K., & Shankland, R. (2005). *An evaluation of end-grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed reading without comprehension, predicting little*. Unpublished manuscript. Literacy Achievement Research Center, Michigan State University. East Lansing, MI

Pugh, K. R., Mencl, W. E., Jenner, A. R., Lee, J. R., Katz, L., Frost, S. J., . . . Shaywitz, B. A. (2001). Neuroimaging studies of reading development and reading disability. *Learning Disabilities Research & Practice, 16*, 240-249.

Pullen, P. C., & Justice, L. M. (2003). Enhancing phonological awareness, print awareness, and oral language skills in preschool children. *Intervention in School and Clinic, 39*(2), 87-98.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). NAEP 2008: Trends in academic progress (NCES 2009-479). Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.

Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation, 14*(13), 1-13.

Ransom, K. A., Santa, C. M., Williams, C. K., & Farstrup, A. E. (1999). High-stakes assessments in reading: A position statement of the International Reading Association. *Journal of Adolescent & Adult Literacy, 43*, 305-312.

Rasinski, T., Blachowicz, C., & Lems, K. (2012). *Fluency instruction: Research-based best practices* (2nd ed.): Guilford

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-Based Measurement Oral Reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469.

Rigney, D. (2010). *The Matthew effect. How advantage begets further advantage*. New York, NY: Columbia University Press.

Roberts, B. (1992). The evolution of the young child's concept of word as a unit of spoken and written language. *Reading Research Quarterly, 27*, 125-138.

Roberts, T. A. (2011). Preschool foundations for reading and writing success. In R. E. O'Connor & P. F. Vadasay (Eds.), *Handbook of reading interventions* (pp. 27-62). New York, NY: Guilford

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third-grade reading comprehension outcomes. *Journal of School Psychology, 46*(3), 343-366.

Rourke, B. P. (1976). Reading retardation in children: developmental lag or deficit? In R. M. Knights & D. J. Bakker (Eds.), *The neuropsychology of learning disorders: theoretical approaches* (pp. 125-137). Baltimore, MD: University Park Press.

Rouse, H. L. F. J. W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*(3), 341-355.

Rumberger, R. W., & Thomas, S. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education, 73*, 39-67.

Safer, N., & Fleischman, S. (2005). How student progress monitoring improves instruction. *Educational Leadership, 62*(5), 81-83.

Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning & Individual Differences, 35*, 103-112. doi: 10.1016/j.lindif.2014.06.009

Samuels, S. J., & Farstrup, A. E. (2006). *What research has to say about fluency instruction*. Newark, DE: International Reading Association.

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A., & Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning and Instruction, 18*(2), 201-210. doi: http://dx.doi.org/10.1016/j.learninstruc.2007.09.017

Scarborough, H. S. (2002). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 1, pp. 97-110). New York, NY: Guilford

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and learning disabilities* (pp. 3-24). Mahwah, NJ: Erlbaum.

Schleicher, A. (2011). Is the sky the limit to education improvement? *Phi Delta Kappan, 93*(2), 58-63.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the reading recovery early intervention. *Journal of Educational Psychology, 97*(2), 257-267.

Seymour, P. H. K., & Evans, H. M. (1994). Levels of phonological awareness and learning to read. *Reading and Writing: An Interdisciplinary Journal, 6*(3), 221-250.

Seymour, P. H. K., & Evans, H. M. (1999). Foundation-level dyslexia: Assessment and treatment. *Journal of Learning Disabilities, 32*(5), 394-405.

Seymour, P. H. K., & MacGregor, C. J. (1984). Developmental dyslexia: A cognitive experimental analysis of phonological, morphemic and visual impairments. *Cognitive Neuropsychology, 1*, 43-83.

Shanahan, T. (2006). The National Reading Panel Report: Practical advice for teachers. Naperville, IL: Learning Point Associates.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., . . . Shaywitz, B. A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3*(1), 69-94.

Shapiro, E. S., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning & Individual Differences, 18*(3), 316-328. doi: 10.1016/j.lindif.2008.03.002

Share, D. L., & Silva, P. A. (1987). Language deficits and specific reading retardation: Cause or effect? *British Journal of Disorders of Communication, 22*, 219-226.

Shavelson, R., McDonnell, L., Oakes, J., & Carrey, N. (1987). Indicator systems for monitoring mathematics and science education. Santa Monica, CA: RAND.

Shaw, R., & Shaw, D. (2002). DIBELS Oral Reading Fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP). *(Technical Report).* Eugene, OR: University of Oregon.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Distribution and temporal stability of dyslexia in an epidemiological sample of 414 children followed longitudinally. *New England Journal of Medicine, 326*, 145-150.

Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*(1), 6-23.

Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences, 23*, 92-100.

Silva, P. A., McGee, R., & Williams, S. (1985). Some characteristics of nine year old boys with general reading backwardness and specific reading retardation. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 26*(3), 407-421. doi: 10.1111/1469-7610.ep11653875

Simmons, D. C., & Kame‘enui, E. J. (1998). *What reading research tells us about children with diverse learning needs: Bases and basics*. Mahwah, NJ: Erlbaum.

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford Press.

Slavin, R. E., Lake, C., Chamber, B., Cheung, A., & Davis, S. (2009). Effective beginning reading programs: A best-evidence synthesis. http://www.bestevidence.org/reading/begin_read/top.htm

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.

Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snyder, L. S., & Downey, D. M. (1991). The language-reading relationship in normal and reading-disabled children. *Journal of Speech, Language and Hearing research, 34*(1), 129-140.

Spear-Swerling, L., & Sternberg, R. L. (1998). *Off track. When poor readers become "learning disabled."*. Boulder, CO: Westview Press.

Spira, E. G., Bracken, S. S., & Fischel, J. E. (2005). Predicting improvement after first-grade reading, difficulties: The effects of oral language, emergent literacy, and behavior skills. *Developmental Psychology, 41*(1), 225-234. doi: 10.1037/0012-1649.411.225

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86*(2), 221-234.

Stanovich, K. E. (1986a). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities* (pp. 87-131). San Diego, CA: Academic Press.

Stanovich, K. E. (1986b). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* Hillsdale, NJ: Erlbaum.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

Tagiuri, R. (1968). *Organizational climate: Explorations of a concept*. Cambridge, MA: Harvard University, Division of Research, Graduate School of Business Administration.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55-64.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7-26.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies *Word recognition in beginning reading* (pp. 148-172). Hillsdale, NJ: Lawrence Erlbaum.

Torgesen, J. K., Rashotte, C. A., & Alexander, A. W. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency and the brain* (pp. 333-355). Timonium, MD: York Press.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 1-15.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology in the Schools, 77*(417-427).

U.S. Department of Education. (2006). The Reading First program's grant application process: Final inspection report. Washington, DC: Office of Inspector General.

U.S. Department of Education. (2010). *ESEA blueprint for reform*. Washington, DC: Author.

U.S. Department of Education, Office of Special Education Programs. (2007). *Final guidance on maintaining, collecting, and reporting racial and ethnic data to the U.S. Department of Education*. (E7-20613).

Uhry, J. K., & Shepherd, M. J. (1997). Teaching phonological recoding to young children with phonological processing deficits: The effect on sight-vocabulary acquisition. *Learning Disability Quarterly, 20,* 104-125.

University of Oregon Center on Teaching and Learning. (2009). Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Retrieved December 23, 2013, 2013, from https://dibels.uoregon.edu/market/assessment/dibels

University of Oregon Center on Teaching and Learning. (2014). UO DIBELS data system. Retrieved October 26, 2014, from https://dibels.uoregon.edu/report/

van Kleeck, A. (1998). Preliteracy domains and stages laying the foundations for beginning reading. *Communication Disorders Quarterly, 20*(1), 33-51.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). The relationship between oral reading fluency and Ohio proficiency testing in reading. Eugene: University of Oregon.

VanDerheyden, A. M. (n.d.). Approaches to RTI. *Selecting an RTI model.* Retrieved October 31, 2014, from http://www.rtinetwork.org/learn/what/approaches-to-rti

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*(1), 2-40.

Vellutino, F. R., & Scanlon, D. M. (1985). Free recall of concrete and abstract words in poor and normal readers. *Journal of experimental child psychology, 39*(2), 363-380.

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of experimental child psychology, 59*(1), 76-123.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1991). Bridging the gap between cognitive and neuropsychological conceptualizations of reading disability. *Learning and Individual Differences, 3*(3), 181-203.

Vellutino, F. R., Scanlon, D. M., & Zhang, H. (2007). Identifying reading disability based on Response to Intervention: Evidence from early intervention research. In S. R. Jimerson, M. K. Burns & A. M. VanDerheyden (Eds.), *The handbook of Response to Intervention: The science and practice of assessment and intervention*. New York, NY: Springer Science.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive test of phonological processes. Austin, TX: PRO-ED.

Walberg, H. J., & Tsai, S. L. (1983). Mathew effects in education. *American Educational Research Journal, 20*, 359-373.

West, J., Denton, K., & Germino-Hausken, E. (2000). America's kindergartners: Findings from the early childhood longitudinal study, kindergarten class of 1998-99: fall 1998 (pp. 126). Washington, DC: National Center for Education Statistics.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 848-872.

Whitehurst, G. J., & Lonigan, C. J. (2001). Emergent literacy: Development from prereaders to readers. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 1, pp. 11-29). New York, NY: Guilford

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement, 28*(1), 61-76.

Willms, J. D. (1994). Monitoring school performance: A guide for educators. *School Effectiveness and School Improvement, 5*(4), 394-397.

Wilson, J. (2005). The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency to performance on Arizona Instrument to Measure Standards (AIMS). Tempe, AZ: Assessment and Evaluation Department, Tempe School District No. 3.

Wolf, M., & Segal, D. (1999). Retrieval rate, accuracy and vocabulary elaboration (RAVE) in reading-impaired children: A pilot intervention programme. *Dyslexia, 5*(1), 1-27.