EXAMINING SCHOOL, STUDENT, AND MEASUREMENT EFFECTS ON

FIRST GRADE STUDENTS' DEMONSTRATION

OF THE ALPHABETIC PRINCIPLE

by

DENI LEE BASARABA

A DISSERTATION

Presented to the Department of Educational Methodology,
Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2011

DISSERTATION APPROVAL PAGE

Student: Deni Lee Basaraba

Title: Examining School, Student, and Measurement Effects on First Grade Students' Demonstration of the Alphabetic Principle

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Educational Methodology, Policy, and Leadership by:

Dr. Paul Yovanoff	Chair
Dr. Gerald Tindal	Member
Dr. Akihito Kamata	Member
Dr. Elizabeth Harn	Outside Member

and

Richard Linton	Vice President for Research and Graduate Studies/Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2011

DISSERTATION ABSTRACT

Deni Lee Basaraba

Doctor of Philosophy

Department of Educational Methodology, Policy, and Leadership

June 2011

Title: Examining School, Student, and Measurement Effects on First Grade Students'
Demonstration of the Alphabetic Principle

Approved: _____
                                    Dr. Paul Yovanoff

Learning to read and successfully decode words is complex, requiring the

integration of critical component skills such as phonological awareness, alphabetic

understanding, and phonological recoding. As foundational skills required for reading

with automaticity, researchers recommend that explicit instruction of these skills begin

early, particularly for students at risk. One commonly used measure to examine students'

alphabetic understanding and phonological recoding skills is DIBELS Nonsense Word

Fluency (NWF), a pseudo-word reading measure composed of vowel-consonant and

consonant-vowel-consonant words.

One purpose of this study was to examine the effects of school-level and individual

student-level predictors on students' overall performance on NWF in the spring of grade 1

as evidenced by their total Correct Letter Sounds (CLS) and Words Read as Whole Units

Correctly (WRWUC) scores. A series of hierarchical linear models were estimated to

investigate the contributions of three student-level predictors (English Learner status, fall of

grade 1 Phoneme Segmentation Fluency raw scores, and fall of grade 1 NWF scores) and

two school-level predictors (percentage of students eligible for free and reduced lunch and

percentage of incoming at-risk kindergarteners) in explaining the variance observed in NWF scores A second purpose was to estimate the item difficulties of the first 20 pseudo-words for comparability of difficulty, controlling for student-level covariates. A series of hierarchical generalized linear models were estimated to investigate the contribution of student-level predictors while controlling for school effects.

Participants were 1,111 first-grade students enrolled in 14 elementary schools participating in the Oregon Reading First initiative. Results indicate that fall of grade 1 NWF raw and quadratic scores were the only statistically significant student-level predictors of CLS and WRWUC scores in the fully specified Level 1 model. The relation between school-level predictors and spring of grade 1 NWF performance complicated interpretation, but both school-level predictors were also significant. Additionally, results of the item difficulty estimates reveal significant student-level effects on item difficulties, providing evidence that item parameters are not equal for the first 20 pseudo-words on DIBELS NWF. The effects were particularly strong for English Learners. Implications for practice and directions for future research are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Deni Lee Basaraba

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

    University of Oregon, Eugene

DEGREES AWARDED:

    Doctor of Philosophy, Educational Methodology, Policy, and Leadership, 2011, University of Oregon
    Master of Arts, Special Education, 2007, University of Oregon
    Bachelor of Arts, Spanish and English, 2003, University of Oregon

AREAS OF SPECIAL INTEREST:

    Schoolwide Beginning Reading Reform
    Assessment and Instructional Decision-Making
    Assessment and Instructional Supports for English Learners
    Curriculum and Instructional Design

PROFESSIONAL EXPERIENCE:

    Supervised College Teaching Assistant, Department of Educational Methodology, Policy, and Leadership, University of Oregon, Eugene, 2010-2011

    Research Assistant, Center on Teaching and Learning, University of Oregon, Eugene, 2003-2011

    Article Reviewer and Coder, National Institute for Direct Instruction, Eugene, OR, 2010

    Project Coordinator, Boys and Girls Club of Portland, Oregon, 2009-2010

    Curriculum Editor, Pacific Institutes for Research, Eugene, OR, 2008

Work Study/Research Assistant, Institute for the Development of Educational Achievement, University of Oregon, Eugene, 1999-2003

GRANTS, AWARDS, AND HONORS:

Inaugural Article of the Year, "They Read but How Well Do They Understand? An Empirical Look at the Nuances of Reading Comprehension," *Assessment for Effective Intervention*, 2011

Dynamic Measurement Group scholarship, University of Oregon, 2010-2011, 2008-2009

Educational Policy Improvement Center scholarship, University of Oregon, 2009-2010

Travel Grant, Department of Educational Methodology, Policy, and Leadership, University of Oregon, 2009 and 2010

Helen DeGnath Wessala scholarship, University of Oregon, 2005-2006

PUBLICATIONS:

Alonzo, J., Basaraba, D., Tindal, G., & Carriveau, R. (2009). They read, but how well do they understand? An empirical look at the nuances of reading comprehension. *Assessment for Effective Intervention, 35*, 34–44.

Smith, J. M., Fien, H., Basaraba, D., & Travers, P. (2009). A culture of continuous improvement: Data-based decision making in a multi-tier system. *Teaching Exceptional Children, 41*(5), 16–22.

ACKNOWLEDGMENTS

To Daniel, for his continual support during many years of graduate school, and to my parents, who always believed I could complete any task I set my mind to.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure   Page

CHAPTER I

STATEMENT OF THE PROBLEM

In this section, I clarify the scope of the problem and propose the provision of early literacy intervention as a method to begin addressing this problem. I then move to a discussion of the importance of providing explicit, systematic, and intensive instruction in phonological awareness and alphabetic understanding, followed by the need to consider student-level characteristics and the need for technically adequate measures when examining students' performance on measures of early literacy. I end with a summary of research questions guiding this proposed study.

The Need for Literacy Instruction

The importance of being a successful, fluent reader cannot be overstated, especially in today's knowledge-driven and technology-driven society. As Adams (1990) and others (National Research Council, 1998; Simmons & Kame'enui, 1998) have argued, not only is the teaching of reading more important than ever to provide students with the myriad of skills needed to succeed in today's rapidly changing, print-driven society, but it is also necessary that this foundational skill be taught better and more broadly than ever. Furthermore, the ability to read is a crucial skill that traverses academic disciplines and provides a sense of personal, social, and economic empowerment for all those fortunate enough to have it (Laugle, 2009; Simmons & Kame'enui, 1998). It is, as Simmons and Kame'enui (1998) also note, "the fulcrum of

academics, the pivotal ability that stabilizes and leverages children's opportunity to learn and to become reflective, independent learners" (p. 1).

Despite widespread recognition of literacy as a critical skill for future success, however, a staggeringly large proportion of children and adults in the United States have demonstrated difficulty learning to read. As of 2007, 42 million American adults were unable to read at all, and another 50 million were unable to read at a level higher than that expected of a fourth or fifth grader; this number increases by approximately 2.25 million each year (National Center for Education Statistics [NCES], 2003). Unfortunately, the figures for students enrolled in today's schools indicate they are not faring particularly well either. Examination of the performance of all students on the fourth-grade reading subtests of the National Assessment of Educational Progress (NAEP) for 2005, 2007, and 2009, for example, reveals that only 24-25% of students have performed well enough to be classified as *Proficient*, while 33-34% have been classified as *Basic* and an alarming 33-36% as having *Below Basic* skills in reading. Only one quarter of the students sampled, in other words, had demonstrated proficiency in beginning reading skills by the fourth grade. Furthermore, examination of student performance data for English learners reveals that this population is not faring well either. In their report for the National Literacy Panel on language-minority youth and children, August and Shanahan (2006) stated that for the 41 states that provided information about the participation and success of English learners on measures of English reading comprehension, only 18.7% of the students who participated in the assessment earned scores above the state-established norm.

The Need for Early Literacy Instruction

These alarming statistics, in addition to a multitude of research, provide support for the provision of explicit, systematic, and intensive instructional supports to students *early,* particularly for students considered to be at risk for later reading difficulties. Results from the National Research Council's (1998) review of the literature led the council to conclude that, "research affirms that quality classroom instruction in kindergarten and the primary grades is the single best weapon against reading failure" (p. 343). This assertion is further supported by the findings of the rigorous review of reading research conducted by the National Reading Panel (National Institute of Child Health and Human Development [NICHD], 2000) and numerous research studies. Findings of the meta-analyses conducted by the National Reading Panel (NICHD, 2000) on the timeliness of phonemic awareness and phonics instruction, for example, revealed statistically larger effects for the provision of instruction in these critical components of reading early (i.e., in kindergarten and first grades) compared to later grades (i.e., in Grades 2-6). Although these larger effect sizes may be due, in part, to the developmental progression of these skills, or the fact that students in the later grades may have already mastered this skill sufficiently so that instruction in these areas is not necessarily useful, findings nonetheless support the provision of instruction early.

Researchers also argue the importance of providing literacy instruction early because of the stability of students' reading trajectories over time (Juel, 1988; Stanovich, 1986; Torgesen, 1998). In his review of the literature, Torgesen (1998) noted that one of the most compelling findings from reading research is that children who begin school with low reading skills and exhibit slow growth in critical word reading skills rarely catch

up to their peers. Torgesen and Burgess (1998), for example, reported that children who experience difficulties with phonological processing skills also experience subsequent difficulties with the rapid, sequential identification and comparisons of sounds, as well as the blending processes that are needed to read words. Furthermore, they presented the findings of longitudinal studies that examined the performance of students from first through fifth grades; these findings indicated that students who exhibited phonological processing difficulties in Grade 1 were not only weak in those skills in Grade 5, but had also experienced relatively persistent reading difficulties across those grade levels. This importance of early intervention, and how it relates to students' reading trajectories over time, has led some researchers (Foorman, Breier, & Fletcher, 2003) to argue that early intervention—particularly in kindergarten and first grade—is more effective than intervention provided in the later grades due to the intensity and duration required; it is more time and cost efficient, in other words, to intervene early.

Phonological Awareness and the Alphabetic Principle:

Critical Components of Early Literacy Instruction

Two skills the National Reading Panel (NICHD, 2000) identified as critical for reading success, particularly for young readers, are phonological awareness and alphabetic understanding. Phonological awareness, or the conscious awareness and understanding that the larger parts of language can be broken down into smaller parts, and phonemic awareness (understanding that the smallest units of spoken language are sounds, or phonemes) in particular are critical skills for learning to read in an alphabetic language (Carnine, Silbert, Kame'enui, Tarver, & Jungjohann, 2006; NICHD, 2000;

Smith, Simmons, & Kame'enui, 1998). This particular type of knowledge is important

for young readers to acquire because these smaller units of spoken language—phonemes,

onsets, rimes, and/or syllables—are blended together to form complete words in English.

 Similarly, awareness and familiarity with these components of spoken language are

useful in decomposing a word into its component parts to understand a word's meaning

(National Research Council, 1998). Moreover, phonological awareness is considered a

critical skill because it helps children develop efficient strategies for recognizing words

by facilitating the storage of sounds in memory that can be easily retrieved when their

corresponding letters are seen in text.

A second skill critical for reading proficiency in English is acquisition of the

alphabetic principle (Adams, 1990; NICHD, 2000; National Research Council, 1998).

The alphabetic principle is comprised of three essential skills: (a) phonological

awareness, or the conscious ability to detect and manipulate sounds in spoken language;

(b) alphabetic understanding, or the understanding that printed letters (or graphemes)

systematically represent the sounds of spoken language; and (c) phonological recoding,

or the ability to blend sounds to read words (Chard, Simmons, & Kame'enui, 1998;

Smith et al., 1998). This second component in particular is critical for learning to read in

languages, such as English, that utilize an alphabetic writing system, but may be difficult

to acquire for two reasons: (a) phonemes are abstract concepts that may be difficult for

young learners to grasp; and (b) there is often no clear, demonstrable relationship

between a grapheme and the sound it represents (National Research Council, 1998). This

is particularly true for English, which is not perfectly alphabetic (i.e., the letters of

English do not have a one-to-one correspondence with the sounds) and can be described

as economical in its attempts to use fewer graphemes to represent a larger number of sounds (Adams, 1990; Juel, 1991). Mastery of these letter-sound correspondences is necessary for students to phonologically recode words accurately and automatically.

Attending to Instructional Design Principles

Although acquisition of phonemic awareness and alphabetic understanding may be difficult for some learners, research suggests that providing students who are struggling to acquire these skills with systematic, explicit, and intensive instruction can be beneficial (Foorman, Breier, & Fletcher, 2003; Foorman & Torgesen, 2001). Instruction that is explicit includes overt and thoughtful explanations of how to complete a task, teacher models of how to complete the task, and guided practice (Rupley, Blair, & Nichols, 2009). When providing explicit phonological awareness instruction, for example, the teacher models for a student how to segment a word into its component parts (e.g., *ssssuuunnn)* and then blends the sounds quickly (e.g., *sun),* leads the students through the task by completing several examples with them, and then tests their understanding. Instruction is considered systematic when it incorporates a thoughtfully planned sequence for content that minimizes the potential for student confusion (e.g., careful attention to the usefulness, and visual and auditory similarity of letter-sounds during beginning phonics instruction) (Carnine et al., 2006). Lastly, the intensity of instruction can be characterized by the inclusion of these two previous principles— explicitness and systematicity in instructional design—as well as by manipulating the format of instructional delivery, specifically time and group size. According to Foorman and Torgesen (2001), increasing the intensity of instruction can be accomplished either

by increasing the amount of instructional time (e.g., from 60 minutes to 90 minutes), or

by decreasing the size of the instructional groups (e.g., from 20 students to five students),

thereby providing students with greater opportunities to practice and receive feedback.

Numerous studies have examined the effects of these instructional design principles on

low-performing students' phonological awareness and phonics skills, leading to findings

that support the utility of these principles for increasing student performance (Denton,

Fletcher, Anthony, & Francis, 2006; Foorman, Francis, Fletcher, Schatschneider, &

Mehta, 1998; Simmons et al., 2007; Wanzek & Vaughn, 2008).

Attending to the Effect of Schools

Although the method of instructional delivery (e.g., explicitness, systematicity,

intensity, etc.) is undoubtedly important, particularly for students at risk for later reading

difficulties (Foorman, Breier, & Fletcher, 2003; Simmons et al., 2007), it is also

important to examine the context within which that instruction is provided—the school—

as research has indicated that schools *do* matter (Baker et al., in press; Fien, Kame'enui,

& Good, 2009). Of particular interest to this study is the understanding of school effects

as measuring the extent of variation *between* schools in the total variation observed in

individual students' test scores (Teddlie, Reynolds, & Sammons, 2000) and, in particular,

an examination of Type A and Type B school effects. According to Raudenbush and

Willms (1995), a Type A effect represents the difference in students' observed

performance on the measures of interest and their expected or predicted performance if

they had attended a "typical school" (in this case, a school that was not implementing a

comprehensive, multitier service model of reading instruction), and a Type B effect takes

into consideration additional school process and school context variables, such as students' socioeconomic status. Examining the effects of school on student performance outcomes is a relevant question because (a) the majority of students receive instruction in a school setting, (b) research has demonstrated that the school students are enrolled in can differentially effect their performance, and (c) recent research has also demonstrated that significant differences in the effect of schools exist even between relatively homogeneous groups of schools, such as those included in this study (Fien et al., 2009).

Attending to Student Characteristics

While research indicates that the role of instruction in contributing to students' success or creating "curriculum casualties" (National Research Council, 1998, p. 25) is substantial, it is also important to consider individual student characteristics when interpreting student performance. Two student-level characteristics are of interest to this study: (a) student's initial skill level on measures of phonological awareness and the alphabetic principle; and (b) students' language proficiency status (i.e., native-English speaker or English learner). Although research indicates that a student's initial skill level is a strong determinant of the growth that he or she will make during the school year (Hedges & Hedberg, 2007), few studies have been conducted examining the contribution of initial skill on later performance (Fien et al., 2008; Good, Baker, & Peyton, 2009). Examining the amount of variance in students' performance accounted for by students' language proficiency status is a practical and critical question due to the surge of English learners (ELs) receiving educational services in schools throughout the United States

(United States Department of Education [USDOE] & National Institute of Child Health and Human Development [NICHD], 2003).

<p style="text-align:center">Attending to the Technical Adequacy of Measures</p>

According to the *Standards for Educational and Psychological Testing (Test Standards;* American Education Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999), test developers have the responsibility to develop tests that permit consistent and valid interpretations of student performance. A review of the literature has revealed that measures of pseudo-word reading, such as DIBELS NWF, are a commonly accepted method for examining students' alphabetic understanding and phonological recoding skills (Chard et al. 1998; NICHD, 2000; Siegel, 1998), and numerous sources of alternate-form and test-retest reliability, as well as convergent and predictive validity, which also speak to the construct validity of the NWF measure (Dynamic Measurement Group, 2008).

Although these numerous sources of validity evidence exist to support appropriate and trustworthy inferences about students' performance on DIBELS NWF from the total scores associated with the measure (e.g., total Correct Letter Sounds), examination of the individual items (i.e., pseudo-words) that contribute to that overall CLS score is lacking. A review of the literature revealed that although researchers have recently begun to examine the equivalency of items on DIBELS ORF via comparison of students' performance on different ORF passages administered over time (Ardoin & Christ, 2008; Francis, et al., 2008), no similar examination of item equivalency for DIBELS NWF has

been undertaken. That is to say, all research conducted thus far with DIBELS NWF has assumed that, based on the facets of word difficulty considered by the test developers, that each pseudo-word is of comparable difficulty for all students although there is not, as of yet, empirical evidence to support this underlying, implicit assumption.

Purpose of This Study

Although findings from numerous studies have demonstrated the importance of phonological awareness and alphabetic understanding to later reading success, as well as the substantial influence that instruction has on these skills, little research has been conducted to purposefully examine the effects that context (i.e., school), student-level characteristics, such as initial skill level and English proficiency, and the measures used to assess students' alphabetic principle skills may have on students' demonstration of their alphabetic principle skills and knowledge. The aim of this study is to address these gaps in our existing knowledge by examining school, student, and measurement effects on first-grade students' demonstration of the alphabetic principle, as measured by students' performance on DIBELS Nonsense Word Fluency (NWF). More specifically, the research questions guiding this study include the following:

1. How much variance in students' performance on DIBELS NWF at both the sound (total Correct Letter Sounds) and word (Words Read as Whole Units Correctly) levels exists between and within schools? How much of that variance is attributable to the following school context variables: (a) the poverty level of the school, and (b) the percent of incoming kindergarten students categorized as being at risk for later reading difficulties?

2. How much of the variance in student performance on DIBELS NWF at both the sound and word levels is accounted for by student-level characteristics, such as student EL status and initial performance on measures of Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF)?

3. Is there significant variation in the item difficulty of the pseudo-words on the spring of Grade 1 NWF benchmark probe? If so, do these items vary significantly as a function of the following possible sources of variance: student-level characteristics (English language status, initial scores on Phoneme Segmentation Fluency, and Nonsense Word Fluency)?

CHAPTER II

REVIEW OF THE LITERATURE

Society is witnessing an explosion in both information and technology, thereby increasing the demand for more sophisticated, print-oriented skills. The social, economic, and personal consequences, however, are dire for those who fail to acquire these critical skills (Adams, 1990; Drucker, as cited in Simmons & Kame'enui, 1998). Recognizing the importance of literacy, policymakers have devoted serious attention and resources to ensuring that the appropriate supports are available to help *all* children learn to read.

National Policy Efforts

The multitude of early literacy research, as well as the amount of money that has been invested in local, state, and federal initiatives focused on early literacy, highlight the understanding that reading is a critical skill necessary for success. On the policy front, for example, the first concerted effort to draw attention to the importance of providing students with the skills they need to be successful was the federal Title I funding of just under $1 billion allocated as part of the Elementary and Secondary Education Act of 1965 (PL 89-10), specifically targeted at improving the academic achievement of disadvantaged children (Sweet, 2004). Since its inception four decades ago, the total cumulative expenditure for Title I has exceeded $140 billion, constituting a substantial investment in the education of children from less fortunate socioeconomic backgrounds found to be at the greatest risk for educational difficulties (Sweet, 2004). In addition to Title I, the stated purpose of which is to ensure that all children, regardless of

12

socioeconomic background, race/ethnicity, or disability status, have a fair and equal opportunity to obtain a high-quality education, several sizable federally funded initiatives have been enacted in the last two decades targeted specifically at helping all students become readers. One such initiative was the Reading Excellence Act of 1998, which provided $260 million to states annually for the establishment of professional development and purchase of instructional materials and assessment instruments aligned with the components of scientifically based reading instruction to teach every child to read by the end of third grade (Reading Excellence Act of 1998; Sweet, 2004).

More recent, and perhaps more controversial, is the Reading First initiative (P. L. 107-110, Part B, Subpart 1), a component of the No Child Left Behind Act of 2001. These pieces of legislation differ in the specificity with which they guide educational initiatives. The ambitious goals of Title I were guided by a series of broad objectives (i.e., closing the achievement gap between high- and low-performing students, aligning high-quality assessments, accountability systems, instructional materials, and teacher preparation, etc.). In contrast, Reading First legislation specified in detail not only the essential components of reading that were the target of the initiative but also provided explicit definitions of critical terms—e.g., scientifically based reading research, screening, diagnostic, and classroom-based assessments—that were the foundational, guiding principles for achieving the goal of every child being able to read at grade level or above no later than third grade (No Child Left Behind Act of 2001, Sec. 1201). Moreover, this piece of legislation was different from previous reading-related initiatives in that it placed the burden of selecting scientifically based instructional materials and diagnostic assessments, as well as providing professional development aligned with these

13

tools, on states and local school districts. According to Sweet (2004), one of the great advantages of this legislation is that "for the first time, the U.S. Congress codified the essential components of reading instruction, confirming the findings of decades of research and writing them into law" (19). Congress's level of commitment to addressing the critical nature of reading skills for all children is further supported by the approximately $15,784,903,000 appropriated for the Reading First initiative from 2002 to 2008 (U.S. Department of Education [USDOE], 2008a). These funds were used to provide intensive support to over 1,800 school districts and more than 5,800 schools located in the United States, District of Columbia, American Samoa, and the U.S. Virgin Islands, including those administered by the Bureau of Indian Education (Gamse et al., 2008). Current estimates indicate that this significant funding was used to provide instructional services and supports for approximately 1.8 million of the nation's students most at risk for not reading at grade level (USDOE, 2008b).

State Policy Efforts

Policy-related efforts to draw attention to the importance of reading have also been made locally, as evidenced by the implementation of the Oregon Reading Initiative (2005-2006) and the recent adoption by the Oregon Department of Education (ODE) of the Oregon K-12 Literacy Framework. The Oregon Reading Initiative of 1999, for example, was supported by federal funds allocated by the ODE and was focused primarily on providing teachers with professional development focused on research-based instructional practices and strategies related to reading (Education Commission of the States, 2002). In December 2009, ODE adopted the Oregon K-12 Literacy

Framework, developed by a committee appointed by the State Superintendent of Public Education, as a tool to facilitate the collaboration of the state, districts, and schools to enable all students to demonstrate proficiency in the Essential Skill of Reading as defined by Oregon diploma requirements. This work was based on the premise that "only a well-coordinated effort that begins in kindergarten, proceeds purposefully through the final year of high school, and involves the active and sustained effort of all levels of the public school system will succeed" (Literacy Leadership State Steering Committee [LLSSC] & Oregon Department of Education [ODE], 2009, p. 8) in achieving the goal of all students becoming proficient, fluent readers. To that end, the Framework provides educators with guidance on how to establish, implement, and support a comprehensive system for reading focused on the following six components: (a) goals, (b) assessment, (c) instruction, (d) leadership, (e) professional development, and (f) commitment.

In addition, the Department has recently provided gratuitous access for all Oregon educators to a multitude of online training modules comprised of presentations, tools, activities, teaching strategies, and video clips of classroom implementation to support teachers' and school leaders' implementation of this comprehensive system for reading (University of Oregon Center on Teaching and Learning and Oregon Department of Education, 2011) Although the Oregon Department of Education is not mandating the adoption of the Framework for schools and districts throughout Oregon, it is demonstrating its commitment to supporting the widespread use of research-based, instructional, systems-level efforts (e.g., classroom, school, district) to increase the literacy knowledge and skills of Oregon's students.

A Wealth of Reading Research

Similarly, the level of commitment to increasing knowledge about the critical components of early literacy has been demonstrated by the research efforts devoted to the topic, as evidenced by the recent publications of reports by the National Research Council (1998) and the National Reading Panel (NICHD, 2000). *Preventing Reading Difficulties in Young Children* (National Research Council, 1998), for example, signaled an effort to end the "reading wars" that have raged in the field of educational research for decades about the best methods available for providing reading instruction (i.e., whole-language vs. phonics-based instruction) by coming to the conclusion that reading should be defined as "a process of getting meaning from print, using knowledge about the written alphabet and about the sound structure of oral language for the purpose of achieving understanding" (National Research Council, 1998, p. 3). More specifically, the members of the committee reviewed a number of informative literatures to establish a series of research-based recommendations regarding the following: (a) promoting literacy development in preschool and kindergarten; (b) the content, timing, and organization of reading instruction; (c) the importance of providing literacy educators with relevant education and professional development opportunities; (d) considerations for providing literacy instruction to English learners; and (e) addressing the needs of students with assiduous reading difficulties.

Furthermore, the report of findings released by the National Reading Panel (NICHD, 2000) provided researchers and practitioners alike with the most comprehensive and rigorous review of reading research literature the field had seen. The

16

panel investigated the contributions of phonemic awareness, phonics, vocabulary, comprehension, and oral reading instruction to the improvement of students' reading skills and how best to provide students across a range of grade levels with adequate instruction. The methodological rigor with which this evaluation was completed far surpassed any attempts to review the literature and summarize the converging findings that had been made previously; all studies included in the meta-analyses for each "big idea" (e.g., phonemic awareness, phonics, fluency with connected text, vocabulary, and reading comprehension) utilized an experimental or quasi-experimental design and provided sufficient details about the sample and methods of the study to allow for the calculation of effect sizes. Overall, this review provided clear and incontrovertible evidence that if teachers were provided with sufficient professional development and training to implement the findings of research, all children could have the opportunity to receive the systematic and explicit reading instruction needed to become successful readers (Sweet, 2004).

The Importance of Early Intervention

Not only has previous experience and research indicated that literacy skills are crucial for students' future success, but research has also indicated that it is critical for students to acquire and have the opportunity to practice literacy skills *early*. In her seminal work, Adams (1990) drew readers' attention to the "catch-22" of early literacy learning, or the fact that the likelihood of a child's success in learning to read is largely dependent on the skills he or she acquired *before* coming to school. That is, children who enter first grade having learned their letters and developed a firm understanding that the

17

words in spoken language are comprised of individual, distinct sounds have already begun learning to read and are likely to become successful and proficient readers.

The National Research Council (1998) arrived at similar conclusions after their review of the literature, although they acknowledged the role of instruction in facilitating students' acquisition of critical skills, claiming, "research affirms that quality classroom instruction in kindergarten and the primary grades is the single best weapon against reading failure" (p. 343). The findings of meta-analyses on phonemic awareness and phonics instruction conducted by the National Reading Panel (NICHD, 2000) also support this conclusion. The National Reading Panel, for example, conducted a meta-analysis to examine the effect of phonemic awareness training on later measures of phonemic awareness and reading performance, using data from 52 studies that met the Panel's rigorous criteria. Examination of the data revealed that the effects of phonemic awareness training in kindergarten (d= 0.95) were significantly larger than the effects observed in first grade (d = 0.48) and in the second through sixth grades (d = 0.70). Although the range of effect sizes was considerably narrower for the relation of phonemic awareness training to reading outcomes (d's from 0.48 to 0.49 for kindergarten through sixth grades), the significant effect sizes nonetheless support the argument for providing instruction early. Moreover, examinations of 38 studies designed to determine the effectiveness of phonics instruction revealed that mean effect sizes of phonics instruction provided in kindergarten and first grades (d = 0.56 and d = 0.54, respectively) were almost twice as large as those obtained in studies in which phonics instruction was provided in second through sixth grades (d = 0.27). Similarly, the Panel found that phonics made significantly larger contributions to younger children's reading growth (d =

0.55) than to older children's reading growth ($d = 0.27$), thereby supporting the premise that instruction in foundational early literacy skills needs to be provided to students sooner rather than later in order to reap the greatest benefits. Although it is possible that the differences in these effect sizes were due, in part, to the developmental nature of reading and the fact that students in the older grades may have already acquired this skill, thereby reducing the effectiveness of instruction on these critical components, the findings nonetheless demonstrate the benefits of providing instruction early.

The results of several seminal research studies also support these findings. Juel (1988), for example, followed the reading and writing development of 54 children as they progressed from first through fourth grades in one elementary school located in a neighborhood with low socioeconomic status. Students participating in the study received reading instruction in a basal series that was categorized as "eclectic" and included some phonics instruction, blending sight words, and the use of context to help identify words. Findings indicated that the majority of children who entered the first grade with poor reading skills were also performing below the grade-level goal at the end of fourth grade. In particular, Juel (1988) found that if children entered the first grade with few phonemic awareness skills at the beginning of the first grade, the probability that they would remain poor readers at the end of fourth grade was .88; conversely, the probability that children beginning first grade with average reading skills would become poor readers was only .12.

Likewise, a comparison of the number of words good and poor readers had seen by the end of the first grade provides additional support for the timeliness of instruction. Juel (1988) found, for example, that while strong first-grade readers had seen, on

average, 18,618 words in connected text by the end of first grade, poor readers had seen, on average, only 9,975 words, or about half as many; the acquisition of critical reading skills early not only increased students' later exposure to words in connected text, but also an increased the number of opportunities to practice their newly learned skills in reading greater numbers of familiar and unfamiliar words.

Furthermore, Foorman, Francis, Shaywitz, Shaywitz, and Fletcher (1997) examined the relation between decoding and reading comprehension in the Connecticut longitudinal data set with students in Grades 1 through 9 to determine when to intervene if a student is experiencing difficulties learning to read. Their findings revealed that the correlations between decoding and comprehension were moderate to strong, ranging from .89 in Grade 1 to .63 in Grade 9 and that between 25% and 36% of the variability in reading comprehension scores on the Woodcock Johnson Psycho-educational Test in Grade 9 were accounted for by early decoding skills. Foorman et al. (1997) concluded that the relation between decoding and comprehension skills was strong and stable over time and also that children's ability to decode words early in their reading education was highly predictive of their ability to comprehend text years later. Although it is possible that this latter finding is a consequence of Matthew effects (Stanovich, 1986), in that students with increased reading skills experience greater exposure to and practice reading text, perhaps the more important finding is that because decoding in the early grades is so strongly related to later reading comprehension skills, reading intervention should begin early and target these critical skills.

Stability of Reading Trajectories Over Time

Research has also supported what has been described a "relatively immutable view of reading" (Foorman et al., 2006), or the idea that, once established, students' reading trajectories are difficult to change (Coyne, Kame'enui, Simmons, & Harn, 2004; Cunningham & Stanovich, 1997; Scarborough, 1998; Stanovich, 1986; Torgesen & Burgess, 1998). In particular, three additional studies examining the impact of early literacy instruction have helped shape the field of literacy instruction and further research. The first of these was a review of the literature conducted by Stanovich (1986), whose purpose was to examine the relation between reading ability and the efficiency of cognitive processes used to read. This review of the literature revealed converging evidence to support not only the idea that students who were better readers were often better decoders—giving them additional opportunities to practice their skills and increase fluency—but also that individual differences in processes such as decoding may cause differences in students' reading efficiency. This latter finding, in combination with the results of earlier studies examining the numbers of words to which good and poor readers were exposed and read in classroom contexts, led Stanovich (1986) to describe these differences as the Matthew effects of reading. Stanovich (1986) used the phenomena described in the story of Matthew from the Bible to describe how strong readers, or the rich, get richer as a result of their strong foundational knowledge and skills, which provide them with increased opportunities to learn, while weaker readers, or the poor, get poorer because they frequently read less due to their frustration and thus experience fewer opportunities to practice their skills and increase their vocabularies.

The findings obtained by Scarborough (1998) in her longitudinal study of students

from the second through the eighth grades also support the idea of the rich get richer and

the poor get poorer, as well as the unfortunate stability of these trajectories over time.

Scarborough (1998) conducted this study with 55 monolingual English speakers with

normal IQs—19 were classified as having a reading disability (RD), while 36 were

classified as not having a reading disability (NRD)—and found that the classifications of

58% of students in the RD group and 97% of students in the NRD group remained stable.

That is to say, the classifications of 55 of the 58 students remained accurate even after 6

years of instruction, indicating that, for the most part, those who were poor readers

remained poor readers, while those who were strong remained strong.

More recently, Torgesen (2000) conducted a review of the literature to examine

the proportion of students who could become proficient, fluent readers as a result of

carefully implemented, explicit and systematic reading interventions. For the purpose of

this review, Torgesen (2000) established the 30th percentile as a normal range of growth

for standard achievement, and the results of his examination of five large-scale

prevention studies led him to three important findings: (a) that once students fall behind

in their growth of critical word reading skills, they may require intensive instructional

interventions to get back on track; (b) losses due to poor reading fluency may be even

more difficult to recoup because of the greatly decreased opportunities struggling readers

have each week, month, or year that they are struggling readers; and (c) in spite of these

facts, the majority of students *can,* with the provision of appropriate instructional

supports*,* become successful, proficient readers. Although the average number of poor

readers for a school population is often cited as between 30% and 60%, as a result of his

review Torgesen (2000) found that all but 2% to 6% of struggling readers can become proficient readers as a result of receiving systematic instruction targeting phonological awareness and decoding skills, particularly when such supports are provided *early*.

Findings from recent studies, however, have arisen to challenge the "relative immutability" of reading but still point to the need to provide reading instruction early (Foorman, Carlson, & Santi, 2007; Phillips, Norris, Osmond, & Maynard, 2002). In a longitudinal study of first- through sixth-grade students with below average, average, and above average reading ability, Philips et al. (2002) found that 30% of students in their sample changed reading categories over time. That is to say, while the trajectories of a strong majority of students (70%) were stable (i.e., below average first-grade students were still below average in sixth grade), it was also possible for some below average and average students to become better readers. Similarly, Foorman et al. (2007) found in their study with 107 first- and second-grade readers that only 31% to 50% of students' reading and spelling achievement was predicted by reading ability at the beginning of the year. Taken together, these findings indicate that while reading trajectories may not be unchangeable, they are typically consistent over time and therefore speak to the need to provide instructional supports to improve performance early.

A Sobering Look at the Current State of Literacy

Recent statistics also indicate, however, that despite preventive efforts, illiteracy rates are still unacceptably high, growth on the National Assessment of Educational Progress (NAEP) has been minimal, and achievement gaps persist between subgroups. According to the results of the National Adult Literacy Survey (NCES, 2003), 42 million

Americans cannot read at all and an additional 50 million are unable to read at a level higher than that which is expected of a fourth or fifth grader. Furthermore, the number of adults who are classified as functionally illiterate increases by approximately 2.25 million per year and about 20% of graduating high school seniors can be classified as functionally illiterate (*Grim Illiteracy Statistics Indicate Americans Have a Reading Problem*, 2007). Recent figures also reveal that 70% of prisoners in state and federal systems can be classified as illiterate, as can 85% of juvenile offenders (*Grim Illiteracy Statistics*, 2007), supporting claims that have been made about the pervasiveness of poor reading ability in our society today that are particularly common to those who are not succeeding in this society (Simmons & Kame'enui, 1998).

One frequently used tool for gauging the literacy health of the nation has been the NAEP, the largest nationally representative and continuous assessment designed to demonstrate what American students know and can do in various subject areas. Recently released statistics indicated that a minimal gain of 4 points was observed for the average reading scale score for fourth graders from 2004 to 2008 and that achievement gaps still exist between subgroups (NCES, 2010b). Females, on average, scored eight points higher than males on the reading subtest of the NAEP, while Caucasian students continued to outperform their African American and Hispanic peers (NCES, 2010c). In addition, the results published in the Nation's Report Card (NCES, 2010d) for 2009 revealed that the average scores for fourth-grade students remained unchanged from 2007 (including for student groups by gender and race/ethnicity), and that only one third (33%) of students performed at or above the level of *Proficient*. Although the authors of the Report Card note that the scores for most student groups were higher in 2009 than when testing began

in 1992, the fact that small gains have been observed over the past 17 years indicates that there is still work to be done.

The outlook for Oregon students is similarly bleak. According to the Alliance for Excellent Education (2009), approximately 11,400 Oregon students failed to graduate from high school in 2009, resulting in a potential loss of lifetime earnings of almost $3.0 billion for the state. This high number of non-graduates may not be surprising in light of the fact that only 34% of all eighth-grade students in Oregon performed at or above *Proficient* on the NAEP in 2009 (NCES, 2010c) and only 69% *met* or *exceeded* standards on the Oregon Assessment of Knowledge and Skills (OAKS), the statewide large-scale assessment in Oregon (ODE, 2009), Similar discrepancies in performance have been observed in the lower grades: 31% of fourth-grade students in Oregon performed *At Proficient* or *At Advanced* on the NAEP (2010c), and 83% of third-grade students *met* or *exceeded* standards on the reading subtests of the OAKS (ODE, 2009).

It is worth noting that achievement gaps exist between groups of Oregon students as well. Examination of 2009 NAEP data, for example, reveals that almost twice as many African American and Hispanic students in Oregon performed at the *Below Basic* level compared to Caucasian students (NCES, 2010c). These same groups were also grossly underrepresented in the *At Proficient* and *At Advanced* levels when compared to their Caucasian and Asian/Pacific Islander peers. Although achievement gaps of this magnitude were not observed in the 2009 OAKS data for third graders, gaps were still evident, with an average of 87% Caucasian students *meeting* or *exceeding* expectations compared to only 68% of Hispanic students and 73% of African American students (ODE, 2009). The fact that these gaps exist as early as third grade and widen as students

25

progress into the upper grades not only provides additional support for the Matthew

effects in reading (Stanovich, 1986), but also supports a call to action for educators:

reading instruction is undoubtedly critical and needs to be provided early.

Theoretical Framework

It is not sufficient, however, that just some form of reading instruction be

provided to students early. On the contrary, research indicates that all students, but

particularly those at risk for later reading difficulties, will benefit the most from

scientifically based reading instruction that provides explicit introduction to and practice

with the skills that have been identified by researchers and practitioners as critical to later

success. Although the focus of this study is on first-grade student acquisition of the

alphabetic principle, the reciprocal relationship between phonological awareness and the

alphabetic principle makes it impossible to discuss one without consideration of the

other. In addition to considering the content of instruction and the importance of the

method of instructional delivery, this study also requires that attention be given to the

context within which this instruction was delivered—the school—for a number of factors

have been found to influence the effectiveness that schools have on improving student

achievement outcomes. Lastly, examination of student performance on a measure

designed to represent their mastery of a complex construct (i.e., the alphabetic principle)

requires the use of a technically adequate measure that not only represents the construct

of interest appropriately but is also designed to function similarly for all students.

Phonological Awareness

Converging evidence, obtained as a result of numerous research studies, supports researchers' claims that phonological awareness is a critical and foundational skill for students learning to read in an alphabetic language such as English (Carnine et al., 2006; National Research Council, 1998). More specifically, phonological awareness has been defined as an understanding and appreciation of the fact that spoken language is comprised of a variety of units, including words, syllables, onsets, and rimes (Adams, 1990; Carnine et al., 2006). In particular, students learning to read in English need to develop *phonemic* awareness skills, a finer-grained sensitivity to language that includes awareness that spoken words can be divided into a sequence of *phonemes*, or individual sounds (National Research Council, 1998). Phonological awareness has also been further defined beyond an awareness of word structure to include the "conscious ability to detect and manipulate (e.g., move, combine, and delete) sounds" (NICHD, 2000; Smith et al., 1998), an ability which is often examined by any of the following tasks (NICHD, 2000; Stahl & Murray, 1998):

1. Phoneme isolation, or the ability to recognize individual sounds in spoken words.

2. Phoneme identification, or the ability to recognize the common sound in different words.

3. Phoneme categorization, or the ability to recognize words with an odd sound in a sequence of odd words.

4. Phoneme blending, or the ability to listen to a sequence of separately spoken

sounds and combine them to form a recognizable word.

5. Phoneme segmentation, or the ability to listen to and deconstruct a spoken word into its individual sounds.

6. Phoneme deletion, or the ability to recognize what word remains when a phoneme is removed.

Moreover, phonological awareness is a learned skill highly related to future reading success. In their review of the literature, for example, the National Research Council (1998) found an extensive research base demonstrating the advantages of providing phonological awareness instruction early, particularly in kindergarten. This finding is not surprising when one considers the abstract physical and psychological nature of phonemes that frequently make acquisition of phonemic awareness difficult for children to acquire spontaneously or independently (Adams, 1990; National Research Council, 1998; NICHD, 2000). Because of this difficulty, many children need explicit, systematic instruction using the types of phonemic awareness and manipulation tasks described above to support their acquisition of this foundational skill.

Furthermore, students who enter school with strong phonological awareness skills or acquire them during the first critical years of their education are more likely to be successful readers than those who do not (O'Shaughnessy & Swanson, 2003; Smith et al., 1998; Torgesen, Wagner, & Rashotte, 1997). O'Shaughnessy and Swanson (2003), for example, evaluated the effectiveness of two different reading intervention approaches with second-grade children identified as having reading disabilities. The interventions under examination were a phonological awareness training (PAT) program, designed to increase students' phonological awareness skills through direct instruction of oral

28

language activities, and a word analogy training (WAT) program designed to increase students' phonological awareness skills through contextualized written language activities. Forty-five children with reading disabilities were randomly assigned to receive six weeks of instruction in either of these reading-related programs or a math-training program. At the end of the study the researchers found that (a) students receiving instruction in the phonological awareness training program made greater gains on a standardized measure of phonological awareness than those students in the word analogy or math training groups, (b) students in both reading intervention programs demonstrated improvement in their knowledge of sound-symbol relations and oral reading fluency skills, and (c) the best predictor of students' growth in speed and accuracy on measures of oral reading fluency was their initial performance on a measure of context-free word reading identification. Although findings (a) and (b) are notable, the last finding provides the greatest support for the importance and contribution of early acquisition of phonological awareness skills, for these findings indicate that the phonological and orthographic skills children *bring* to reading instruction that are the most important determinants of their ability to benefit from that intervention (O'Shaughnessy & Swanson, 2003).

Alphabetic Principle

Just as researchers agree that understanding of the sound structure is necessary for being able to read in an alphabetic language such as English, so too have they concluded that it is critical for students to acquire the understanding of a systematic relation between the phonemes of spoken language and the graphemes printed on the page that represent

those sounds (Adams, 1990; Juel, 1991; National Research Council, 1998). Researchers

(Fien et al., 2010; Smith et al., 1998) have defined the alphabetic principle as a larger

construct that encompasses three skills critical for reading success: (a) phonological

awareness, or the understanding that words are comprised of individual sounds, or

phonemes; (b) alphabetic understanding, or the understanding that printed letters

systematically represent those sounds and that whole words have a sound structure that

consists of individual sounds and patterns of groups of sounds; and (c) phonological

recoding, or the ability to blend sounds to read words. These latter components,

alphabetic understanding, and phonological recoding, are frequently examined through

students' knowledge of letter-sound correspondences, or their accuracy in identifying the

linkages between discrete phonemes and individual letters, or graphemes (Adams, 1990;

Smith et al., 1998).

One criticism of focusing on solely letter-sound correspondences (independent of

word and text reading) is that it may be akin to teaching in a vacuum by denying students

the opportunity to apply their newfound understanding of print to real words and

connected text (Chard et al., 1998). As Adams (1990) noted, learning these individual

letter-sound correspondences and generalizations of phonics rules are "inherently

tractable when divorced from the rest of the reading situation. They are abstract,

piecewise, unorderable, unreliable, barely numerable, and sometimes mutually

incompatible" (p. 291), thereby requiring multiple opportunities to practice across

multiple contexts. It is important to remember, however, not only that letter-sound

correspondences are intended only as an *initial* step in acquisition of the alphabetic

principle, but also that providing students the opportunity to practice their newfound

30

knowledge in this particular context is critical because it provides them with information about how to connect familiar sounds with symbols they know are important for making meaning in text (Chard et al., 1998). Moreover, the National Research Council (1998) concluded that there is no need to wait until children know all of the letters of the alphabet (and their corresponding sounds) to begin providing explicit decoding instruction, but rather that simple decoding instruction (i.e., with VC and CVC words containing familiar sounds) can begin as soon as students have mastered the values of several consonants and vowels. The importance of acquiring the alphabetic principle is emphasized by findings that indicate that a failure to acquire this skill is one of the main stumbling blocks known to throw students off course on their journey to becoming skilled readers (National Research Council, 1998).

Relation Between Phonological Awareness and the Alphabetic Principle

Because the relation between phonological awareness and the alphabetic principle is a reciprocal one (Adams, 1990; Liberman, Shankweiler, & Liberman, 1989; Perfetti, 1985; Stanovich, 1986), it is almost impossible to discuss one without referring to the other. The reciprocity between these two concepts is due, in part, to the alphabetic structure of the English language that is represented by phonemes and graphemes; understanding that words are comprised of individual sounds is just as critical as the ability to be able to link those sounds to letters (Perfetti, 1985). According to Adams (1990), "functional understanding of the alphabetic principle depends equally on knowledge of letters and on explicit awareness of phonemes because it depends integrally on the association *between* [emphasis added] them" (p. 304). Furthermore, not only does

knowledge of phonemes facilitate the ability to correctly link letters to sounds, explicit instruction regarding the correspondences between letters and sounds facilitates growth in phonological awareness (National Research Council, 1998). Additionally, support for this bidirectional, reciprocal relation between phonological awareness and the alphabetic principle can be found in two types of studies, those that have demonstrated that phonological awareness by itself is an insufficient condition for learning to read and those in which better student outcomes on measures of reading were obtained when phonological awareness instruction was paired with letters.

Phonological Awareness Alone Is Insufficient

Because successful reading requires the ability to correctly map sounds onto the letters that appear on the printed page, it may not be surprising that simply having knowledge of the sounds is an insufficient condition for learning to read (Bus & Van Ijzendoorn, 1999; Fielding-Barnsley, 1997; Foorman, Chen, et al., 2003). Bus and Van Ijzendoorn (1999), for example, conducted a meta-analysis of studies conducted both in the United States and Europe to determine if phonological awareness training was more effective when (a) combined with written letters or words, and (b) provided earlier rather than later in a child's education. To answer these questions, the researchers reviewed 62 published studies that utilized measures of phonological awareness and/or reading as their outcome measures. These studies were conducted in a number of different countries (i.e., European nations and the United States), and with populations varying by student skill level, age range, and school setting. Of the studies conducted in European countries, the results indicated that a purely phonetic training was less effective (combined *d* =

1.19) than phonological awareness training that included letters (combined $d = 1.75$).

Interestingly, these same findings were not obtained for studies conducted in the United

States; rather, purely phonetic programs were found to be as effective as those that were

embedded in letter-sound training or reading and writing practice. Bus and van

Ijzendoorn (1999) hypothesized that the reason for the inconsistent findings may be

because purely phonetic programs, such as those implemented in the European studies,

are rarely implemented in the United States.

Foorman, Chen, et al. (2003) obtained similar findings in their review of assorted

published reading curricula to determine the characteristics of effective phonological

awareness instruction and whether that instruction should be solely auditory in nature or

provided in the context of letters. During this study, Foorman, Chen, et al. (2003)

examined the performance of kindergarten students enrolled in schools that had received

Title I funding and where teachers implemented a variety of reading curricula to

determine if the amount of phonological awareness instruction in a program (i.e., more or

less explicit phonological instruction) and the amount of choice a teacher had in

providing literacy instruction had any effect on student performance. They found that

alphabetic instruction without phonological awareness instruction was not as effective as

alphabetic instruction *with* phonological awareness instruction. More specifically, results

indicated that activities in which phonemes were first blended and segmented in speech

and then explicitly and systematically connected to graphemes during phonics instruction

were the most beneficial for students. These findings not only support the integration of

phonological awareness and phonics instruction but also the claim that phonological

awareness instruction alone is insufficient for improving students' reading skills.

Pairing Phonological Awareness Instruction With Print

Although phonological awareness is primarily an auditory measure (Smith et al., 1998), phonological awareness interventions appear to be more effective in helping children learn to read when paired with instruction using letters. According to the National Reading Panel (NICHD, 2000), one of the reasons phonological awareness is so important is because it helps children grasp how the alphabetic system works in their language, and therefore it may make sense for teachers to use letters during phonological awareness instruction. Torgesen et al. (1999) purposefully examined the advantages of this instructional approach with 138 students who began kindergarten with delayed phonological awareness skills. More specifically, the effectiveness of three different intervention programs was examined. They compared (a) an explicit phonological awareness intervention in which time was spent on phonological awareness, phonemic decoding, and reading decodable text (PASP); (b) an embedded phonics intervention in which students played word-level drill games, learned to read words in context, and practiced reading and writing sentences (EP); and (c) regular classroom reading instruction, which consisted of a variety of phonics-oriented activities (RCS). Students were randomly assigned to one of the intervention groups and received four 20-minute sessions of 1:1 instruction for 2½ years. At the end of second grade, students who participated in the PASP intervention had significantly stronger phonemic awareness, phonemic decoding, and context-free word reading skills than students who had received instruction in the EP and RCS groups. Although no causal link between student progress and instructional practices was established, it is worth noting that the researchers found

that teachers in the PASP condition spent almost twice as much time on word-level instruction than text-level instruction (80% compared to 43%), which may account for the differences observed in student performance.

Similar findings were reported by Torgesen et al. (2001), who studied the progress of 60 children between the ages of 8 and 10 to compare the effectiveness of two approaches that contained explicit instruction in word-level skills but varied systematically in their depth of phonemic awareness and extent of practice provided in decontextualized phonemic decoding skills. One program used during the study was the Auditory Discrimination in Depth (ADD) program, which focused on teaching students the kinesthetic, auditory, and visual (mouth) form features associated with the production of all of the common phonemes in the English language and, as they learned to label those phonemes, also taught students how to associate a specific letter with the phoneme and used small plastic tiles (i.e., manipulatives) to learn to encode (spell) and decode (read) words. The second program used was the Embedded Phonics (EP) program, during which students received direct, explicit instruction in word reading skills and numerous opportunities to practice reading and writing meaningful text. Students received approximately 68 hours of 1:1 instruction in one of the aforementioned interventions. Students who received instruction in the more explicit ADD program demonstrated higher growth rates on measures of decoding and accuracy and fluency with connected text than those who received instruction in the less explicit EP program.

Challenges Posed by the Alphabetic Writing System of English

Although studies have indicated that explicit instruction in phonological awareness, when paired with letters, can help students who are struggling to become proficient, fluent readers, one challenge consistently associated with acquisition of the alphabetic principle in English is that it relies on an alphabetic writing system (Adams, 1990; Juel, 1991; National Research Council, 1998; Perfetti, 1985). This alphabetic writing system poses a challenge because the units represented graphically on the page are phonologically abstract and referentially meaningless; there is no explicit reason that the letter "b," for example, represents the sound /b/ and not the sound /p/ (Lyon, 2009; National Research Council, 1998). Furthermore, although letters map onto phonemes and phonemes map roughly onto perceptually and functionally relevant targets, there is not always a 1:1 correspondence between a sound and the letter(s) used to represent it (Adams, 1990). As Lyon (2009) and others (Adams, 1990; NICHD, 2000) have noted, for example, students need to learn how to use the 26 letters of the English alphabet to represent the approximately 40-44 phonemes of the English language.

This complex relationship is the result of the economy of the alphabetic writing system of English—allowing almost direct phoneme-grapheme relationships for consonants but complex correspondences for vowels (Adams, 1990; Juel, 1991). English, for example, has only five standard vowel letters, but approximately 12-15 vowel sounds (Frost, 2005; Juel, 1991). Juel (1991), in fact, asserts that these orthographic complications that are the result of such economy are also the primary obstacles to learning the alphabetic principle for beginning readers: (a) the abstract nature of

phonemes, consonants in particular; and (b) the failure of the writing system to use a distinct letter to code each vowel.

The complications associated with its complex orthography have led to the characterization of English as a language with a deep, opaque, or even "irregular" orthography (Cardoso-Martins, 2001; Foorman, Breier, & Fletcher, 2003; Frost, 2005; Genesee, Geva, Dressler, & Kamil, 2006). As noted earlier, the English language lacks a 1:1 correspondence between the number of sounds in the language and the numbers of letters used to represent those sounds. Although consonantal mappings are fairly consistent, there are exceptions, such as the sound /f/, which can be represented in English by three graphemes: "f" as in *find*, "ph" as in *philosophy,* and "gh" as in *tough* (Genesee et al., 2006). Furthermore, English words rely on a rich vowel system of approximately 15 vowels, which are represented by only five graphemes.

According to Frost and others (Plaut, McClelland, Seidenberg, & Patterson, 1996), the complexity of grapheme-phoneme correspondences is heavily influenced by the transparent or opaque mapping of spelling patterns, which is in turn influenced by two distinct factors: regularity and consistency. Regularity is defined as the conformity of clusters of letters (e.g., *ch*) to grapheme-phoneme correspondence rules, while consistency refers to the uniqueness of the pronunciation of words. Examples of irregularity include the words *yacht* and *chef,* neither of which can be computed using grapheme-phoneme correspondence rules. The words *moth* and *both*, on the other hand, are examples of the lack of consistency in the English language, for while both words contain the same orthographic body (*-oth*) they are not pronounced the same. It is possible, in other words, for words in the English language to be regular but inconsistent

(e.g., *moth* and *both*), or irregular and consistent (e.g., *kind, bold,* or *took;* Plaut et al., 1996), and because English contains many words that are either irregular or inconsistent, it is characterized as a language with deep orthography.

The deep orthography of English has frequently been posited as a factor that complicates the application of an alphabetic reading strategy, particularly for struggling readers (Cardoso-Martins, 2001). According to Dressler and Kamil (2006), this may be due to the script-dependent hypothesis that emphasizes the importance of the nature of orthography in learning to read. In particular, this hypothesis suggests that the process of mapping graphemes to their corresponding phonemes when reading is easier in orthographic systems that are characterized by regular sound-symbol relationships and, thus, are relatively transparent.

This is not to say, however, that word recognition via an alphabetic reading strategy in English is impossible, for as Foorman, Chen, and Fletcher (2003) observed, approximately 50% of spellings in English follow grapheme-phoneme correspondence rules, and another 36% follow these same rules with only one error. This means that approximately 14% of English words contain "irregular" spellings, a percentage that decreases to only 4% once one considers word origin, meaning, and morphology. These findings in particular led Foorman, Chen, et al. (2003) to conclude that the primary objective of beginning reading instruction should be to help children master the alphabetic system (letter-sound correspondences and grapheme-phoneme correspondence rules) for the 86% of words that follow the rules, to use other linguistic cues to help them read the 10% of words whose regularity is based on word origin, meaning, and morphology, and to memorize the 4% of words that are true oddities. Nevertheless, it is

likely that all students, but particularly those at risk for later reading difficulties, will benefit from systematic instruction that explicitly attends to not only the grapheme-phoneme correspondences needed to decipher the majority of English rules but also to the other rules and relationships needed to decode the remaining words.

Acquisition of the alphabetic principle is also complicated by the fact not all sounds in a spoken word receive equal attention and can be distinguished and heard. Adams (1990) pointed out, for example, that many sounds—consonants in particular—cannot be pronounced in isolation; although we can say /puh/ or /pah/ for "p," pronouncing the /p/ by itself is virtually impossible. Similarly, the National Research Council (1998) noted that although the word "but" consists of three sounds that are technically represented by three letters (/b/ /u/ /t/), each sound in isolation does not refer to anything and only the medial sound /u/ can be pronounced by itself. This fact, in combination with the co-articulation of sounds in spoken language (i.e., sounds are not readily distinguishable from one another in spoken speech) has led reading researchers and designers of reading instruction texts alike to establish guidelines for letter-sound order when teaching students how to phonologically recode, or blend the individual letters of a word together.

Adams (1990) and Siegel (1998), for example, have advocated for the introduction of consonants before vowels because consonants, with "gratifying frequency" (Adams, 1990), tend to have a 1:1 correspondence between phonemes and the graphemes used to represent them, unlike vowels, which have been described as "rampantly irregular" (Adams, 1990) due to the complex representation discussed earlier. Attention needs to be paid, however, to the order in which consonants are introduced,

relative to their position in words because how they are pronounced affects students'

ability to learn decoding skills (Kame'enui & Simmons, 1990). More specifically,

consonantal sounds in the English language fall into two broad categories: stop sounds

and continuous sounds. Stop sounds, or plosives (/b/, /p/, /d/, /t/, /k/, /g/), are formed

when air is built up in the vocal tract and suddenly released through the mouth. During

the production of stop sounds, in other words, the airway is completely obstructed and a

rush of air is released once the obstruction is opened (Finegan, 2008). Because of the

sudden release of air required to produce the sound, stop sounds cannot be elongated, or

stretched. Continuous sounds, in contrast, allow for prolonged stretching of the sound

because the airway is not completely obstructed and the positioning of the tongue and lips

allows for a continuous release of air to sustain the pronunciation of the sound (Finegan,

2008; Kame'enui & Simmons, 1990). Because continuous sounds are easier to hear and

distinguish, can be elongated, and allow the reader to transition from one sound to the

next, instructional design experts agree that beginning reading instruction—and decoding

instruction in particular—ought to start with words that begin with continuous sounds

(Chard & Osborn, 1999; Juel, 1991; Kame'enui & Simmons, 1990; Smith et al., 1998).

<div align="center">Paucity of Research Related to Letter-Sound Production and Reading</div>

Although much literature exists regarding the complications presented to young

readers by the difficulties of the alphabetic writing system of English and the majority of

instructional texts suggest that continuous sounds ought to be introduced before stop

sounds to facilitate the blending of sounds into words, one area in which surprisingly

little research has been conducted is the relation between letter-sound production and the

ability to blend sounds into words. Furthermore, what research does exist presents

conflicting findings that may simply confuse educators. Treiman, Tincoff, Rodriguez,

Mouzaki, and Francis (1998), for example, investigated the pronounceability and

syllable-position hypotheses with students ranging in age from 3.5 to 7.5 years by

examining to what degree their knowledge of letter names facilitated their knowledge of

letter sounds. According to the researchers (Treiman et al., 1998), the pronounceability

hypothesis is based on the premise that stop consonants, which cannot be pronounced

without vowels, are difficult for students to identify as separate units, while the syllable

position hypothesis claims that obstruent consonants (e.g., /s/, /p/, /b/), which are

typically found at the beginnings and ends of syllables, are easier than sonorant

consonants (e.g., /w/, /m/, /r/), which can appear in the middle of syllables, for children to

learn (Treiman et al., 1998).

The researchers (Treiman, et al., 1998) found that while one of the three data sets

analyzed produced results that supported the syllable position hypothesis, the other data

sets did not support this hypothesis, prompting the researchers to claim that there is no

consistent evidence to support the syllable position hypothesis. In contrast, regression

analyses of these three data sets did not reveal any significant differences in the difficulty

of stop versus sonorant (i.e., continuous) consonant production, allowing the researchers

to suggest that the properties of the phoneme do not appear to have a consistent influence

on children's ability to map the sound to print (Treiman, et al., 1998). Although these

findings are compelling and relevant to this study, it is important to note that these claims

were made on the basis of sound production in isolation. That is to say, students in this

study were asked to identify letter names and letter sounds when provided with examples

of them in isolation and were not asked to blend any of the sounds together as they are required to do when reading words.

Perhaps it is this critical distinction that has led authors of literacy instruction textbooks to vastly different conclusions about the importance of the phonological properties of sounds in learning to read. According to Smith et al. (1998), not only are phonemes in words co-articulated and therefore subject to distortion, but one of the factors that contributes to the confusability of phonemes and difficulty in production are the phonological properties of sounds (e.g., continuant sounds, such as /m/, are easier for students to produce and segment than stop sounds, such as /t/). Furthermore, these researchers note that in the nine studies they were able to locate that considered the relative difficulty of phoneme position in words,

> continuant sounds are typically introduced before stop sounds, because stop sounds are more difficult to elongate and, therefore, more difficult to isolate, detect, and manipulate. Stop sounds were often introduced later because of the articulatory distortion that often occurs when a stop sound is produced in isolation. For example, it is difficult for many children to detach the vowel sound /u/ when voicing the /t/. (Smith et al., 1998)

Closer examination of these studies (Ball & Blachman, 1991; Byrne & Fielding-Barnsley, 1989; Cunningham, 1990; Defior & Tudela, 1994; Lie, 1991; Lundberg, Frost, & Petersen, 1988; O'Connor, Jenkins, & Slocum, 1995), however, revealed that the relative difficulty of phoneme position in words was not the topic of interest and therefore never explicitly examined by these researchers. Although researchers attended to the order of phoneme introduction in the design of their respective interventions, that attention to sound order was based on a common but empirically unproven belief that continuous sounds ought to be introduced before stop sounds during beginning reading

instruction in order to facilitate students' acquisition of blending skills. Allor and

McCathren (2003), who perpetuated this belief in their description for teachers of

strategies that can be used with storybooks to facilitate early literacy, make a similar

recommendation: that teachers first introduce words that contain continuous initial

phonemes, such as /s/ or /f/, which can be stretched, as opposed to stop sounds, which

may be more difficult to blend. Because there is, as yet, little empirical data regarding the

role of sound production in blending sounds to read words, one potential area for future

research is whether the difficulty of simple CVC words or pseudo-words commonly used

to gauge students' decoding skills is influenced by the type of sound (e.g., continuous

versus stop sound) that each word begins with.

Instructional Supports for Students at Risk

Although acquisition of phonological awareness and the alphabetic principle is

critical for later reading success, equally important for students, particularly those at risk

for reading difficulties, are the instructional supports provided to facilitate the acquisition

of those skills. Instruction for children who are experiencing difficulties learning to read

must be more systematic, explicit, intensive, and supportive than the instruction required

for the majority of children (Carnine et al., 2006; Foorman & Torgesen, 2001; Foorman,

Breier, et al., 2003; Rupley et al., 2009; Stahl, Duffy-Hester, & Dougherty Stahl, 1998).

Benefits of Explicit Instruction

In their textbook designed for educators, Carnine et al. (2006) provide readers

with a series of guidelines to help them clearly, overtly, and thoroughly communicate

information to students. The components of explicit, systematic instruction include (a) a model, or demonstration by the teacher for the students on how to complete the task (e.g., pointing to the letters in the word while producing their sounds, */m/ /a/ /n/*, and then blending the sounds *mmmaaannn*); (b) a lead step in which the teacher helps students produce the desired response (e.g., by making the response with students as they try to apply the skill); and (c) a test in which the students complete the task without teacher support (Carnine et al., 2006). For phonics instruction specifically, this explicit approach might include the presentation of a written letter on the board ("m") and a teacher modeling for students the production of the sound, "/mmmmm/." Once the teacher has provided several models, the teacher then guides students in their pronunciation of the sound by pointing to the letter on the board and having them produce the sound along with him/her, following this guided practice with individual turns to test students' mastery of the letter-sound correspondence (Chard et al., 1998). In this initial instruction context, all letter-sound pairs are presented in isolation, without the distraction of other correspondences to ensure that the letter-sound correspondence is salient and, with multiple opportunities to practice, automatic.

The provision of explicit instruction significantly benefits struggling students (Blachman, Tangel, Ball, Black, & McGraw, 1999; Foorman et al., 1998; McClandiss, Beck, Sandack, & Perfetti, 2003). Foorman et al. (1998), for example, conducted a study with 285 first and second graders who were eligible to receive Title I services in 19 elementary schools to test their hypothesis that children who received explicit instruction in the alphabetic principle with a focus on letter-sound correspondences would show greater growth compared to students receiving less explicit instruction. To test this

hypothesis, the researchers randomly assigned children to one of three groups: (a) a direct code (DC) intervention program that utilized a balanced emphasis on phonemic awareness, phonics, and literature-based activities; (b) an embedded code (EC) intervention program that provided emphasis on phonemic awareness and spelling patterns in the context of predictable text; and (c) an implicit code (IC) intervention that provided students with teacher facilitated instruction in a print-rich environment. Student performance on a battery of reading measures after a yearlong intervention provided three sources of evidence supporting researchers' advocacy for explicit instruction. Students who participated in the DC intervention and received direct, explicit phonics instruction improved in word-reading skills significantly faster than their peers in other interventions (46% of students in the IC and 44% of students in the EC intervention showed no demonstrable growth in word reading, compared to only 16% of the DC group). Additionally, not only did students in the DC approach national averages on measures of decoding and passage comprehension while students in the other groups did not, but those with initially low phonological processing skills also showed significant growth in pseudo-reading skills.

Similar findings supporting the effectiveness of explicit instruction have also been obtained for students in the lower grades. Blachman et al. (1999), for example, conducted a 2-year longitudinal study with kindergarten and first-grade students enrolled in demographically comparable, low-income, inner-city schools in a large urban district. During the first year of the study, the 84 kindergarten students assigned to the treatment condition received 15-20 minutes of explicit phonological awareness instruction in small groups of four to five students for 41 intervention sessions. At the end of the school year,

45

students who received the intervention significantly outperformed their peers in the control group on measures of phonemic awareness, letter name and letter sound knowledge, and reading phonetically regular real and pseudo-words. During the second year of the study, students were rank-ordered based on their letter name and sound knowledge, phoneme segmentation skills, and word reading skills and separated into small groups. The reading intervention during the second year consisted of phonemic awareness and letter sound review, phoneme analysis and manipulation instruction, decoding and high frequency word instruction and connected text reading. At the end of the intervention, students in the treatment group performed significantly better than those in the control group on measures of letter names, letter sounds, real word reading, pseudo-word reading, and spelling, thereby providing additional evidence for the advantages of explicit instruction for students considered to be at risk for later reading difficulties.

Strategies for Intensifying Instruction

Foorman and Torgesen (2001) noted that there are two ways to intensify instruction to better support students at risk: increasing instructional time (e.g., from 90 minutes to 120 minutes) and/or providing instruction in small groups. Although increasing instructional time is an obvious way to intensify instruction, providing instruction in small groups is another useful method because it provides students with more opportunities to practice their newly learned skills and receive teacher feedback on their performance. Research has not only indicated that students benefit from opportunities to practice skills repeatedly (even to the extent of overlearning) and in new

contexts (Adams, 1990; Rupley et al., 2009), but also that specific and prompt teacher feedback increases students' learning and motivation to learn (Clifford, 1990). Attending to other instructional variables such as pacing (Carnine et al., 2006), scaffolding (Coyne, Kame'enui, & Simmons, 2001; Foorman & Torgesen, 2001; Kame'enui & Simmons, 1990), developing a thoughtfully-planned sequence for delivery of the content (Chard et al., 1998), and providing judicious review (Coyne et al., 2001) have also been found effective for students at risk for later reading difficulties.

Establishing a Connection Between Instruction and Word Reading

Various theories of reading, and word reading specifically, may provide insights into the types of instructional supports students need to acquire alphabetic understanding and to be able to read new words. Most relevant to this study, however, are the phases of word reading development proposed and refined by Ehri (1992; 1997; 2005b), as recent research based on this theory has demonstrated the importance of phonological recoding, or reading words as whole units, to later reading performance.

Ehri's Phases of Word Reading Development

Ehri situates her theory of word reading within the context of four general approaches to word reading: decoding (or phonological recoding); analogizing, predicting, or memorizing. The phases of word reading development illustrated in Figure 1 are most readily applicable to decoding, or the process of sounding out and blending graphemes into phonemes.
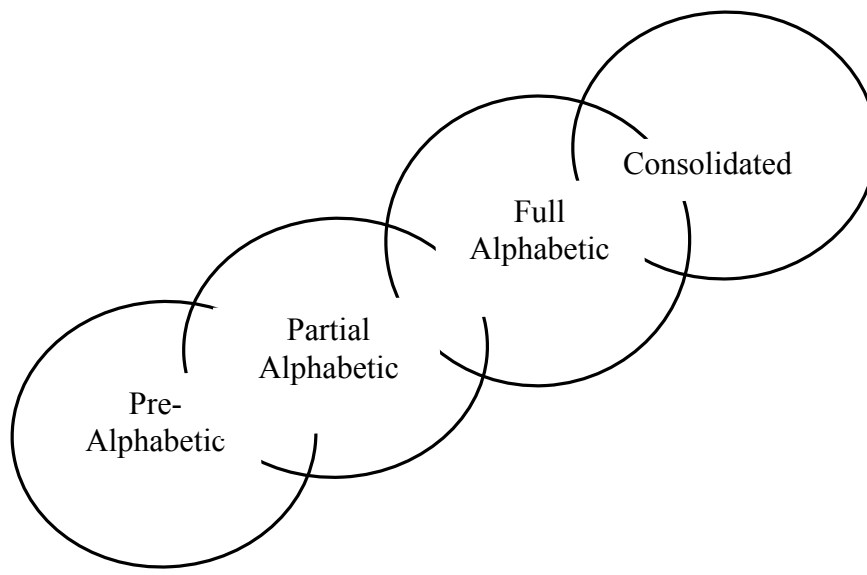
FIGURE 1. An illustration of Ehri's (2005a) phases of word reading
development.

This theoretical framework proposed by Ehri (2005b) and others (Ehri, 1999; Ehri

& McCormick, 1998; Ehri & Snowling, 2005; Perfetti, 1999) introduces the possibility

that students progress through four phases of development when learning to decode

words: pre-alphabetic, partial alphabetic, full alphabetic, and consolidated alphabetic

phases. During the pre-alphabetic phase, children rely primarily on environmental cues

(as opposed to alphabetic knowledge) to read words, as they have little understanding that

the letters in written words systematically map onto the sounds they hear in spoken

language. Once children have acquired this understanding, have learned the sounds of

letters in the alphabet, and use this knowledge to remember how to read words, they have

progressed to the partial alphabetic phase. As one might expect, because students in this

phase lack full knowledge of the alphabetic system and thus continue to experience

difficulty with some letter-sound correspondences (especially vowels), word reading

during this phase is an imperfect process potentially rife with errors for those students'

whose letter-sound correspondence knowledge is not firm (Ehri, 2005a).

Progress to the full alphabetic phase occurs when children are able to form

complete connections between graphemes and phonemes in pronunciations and are able

to segment words into phonemes that match up to the graphemes they see in a printed

word. As children retain more sight words in their memory (i.e., words that can be

automatically accessed because students have firm understanding of the relation between

the phonemes and graphemes), they progress to the consolidated phase in which

grapheme-phoneme connections in words are stored in memory as larger units. Although

each of these phases has been described here in a sequential, linear fashion, it is

important to note that one of the reasons Ehri (2005b) and others (Harn, Stoolmiller, &

Chard, 2008; Seymour, 2005) use the term "phases" instead of "stages" is because

mastering the skills in one phase is not a prerequisite for progression to the next phase; it

is possible, rather, for students to demonstrate characteristics of multiple word reading

phases as they approach unfamiliar words.

<center>Examining the Role of Automaticity in Word Reading</center>

One critical skill associated with fluent decoding, or the ability to blend sounds

together effortlessly and accurately to read unfamiliar words, is automaticity (Adams,

1990; Juel, 1991). The importance of automaticity in the development of reading skills

was first proposed by LaBerge and Samuels (1974), who argued that mastery of the

component skills of reading was necessary to be automatic and fluent readers. More

recently, automaticity has been defined as "the ability to perform complex skills with

<center>49</center>

minimal attention and conscious effort" (Samuels & Flor, 1997). In their initial discussion of automaticity and how it might be studied, LaBerge and Samuels (1974) provided an apt example relevant to this discussion of word reading, for they noted that two criteria of achievement are applicable in learning to sound letter patterns: accuracy and automaticity. To achieve accuracy, for example, a reader must have the ability to form direct associations between the phonemes of spoken language and the graphemes that are printed on the page. Although accuracy in letter-sound correspondences is important in learning to read, by itself, it is insufficient. Readers must also be able to blend sounds into syllables or words, which is where automaticity comes into play. If a student is not firm on his or her letter-sound correspondences and needs to devote a good deal of effort to accurately identifying letter-sounds, blending will be more difficult because of the amount of information (i.e., letter-sound correspondences) that needs to be attended to and held in short-term memory. Conversely, if students have mastered letter-sound correspondences they can devote a greater amount of attention to blending those sounds into words, conditions under which decoding requires minimal attention and effort.

More recent conceptualizations of the role of automaticity in reading are a bit more complex. Logan (1997), for example, posited that automaticity is comprised of four components: (a) speed, (b) effortlessness, (c) autonomy, and (d) lack of conscious awareness. Within this framework, speed is a critical component because an increase in speed is directly associated with a decrease in reaction time. Increased familiarity with letter-sound correspondences, for example, means that a student is able to identify them more rapidly and require less "think time" to correctly match a phoneme with the

grapheme used to represent it. Additionally, tasks that have become automatic require minimal (if any) effort and, as Logan (1997) noted, the effortlessness of tasks that have become automatic is first evident as a sense of ease and later evident in a person's ability to do another task while performing an automatic one. In learning to read, for example, one might argue that when students have demonstrated their ability to decode unfamiliar words accurately and fluently they are demonstrating the automaticity of their letter-sound correspondence knowledge; no longer do they have to attend to each letter individually but rather are able to blend the letters of a word together easily because their ability to map sounds to print has become automatic.

Furthermore, when readers acquire automaticity with letter-sound correspondences (or any other reading-related skill, for that matter) the cognitive processing skills required to complete the task are autonomous, no longer requiring the readers' conscious and deliberate attention. On a related note, once automaticity has been acquired, conscious, deliberate attention is no longer required to complete the task in question and, in fact, is not even available to our consciousness. Once we have mastered the component skills of reading, for example, it is difficult to consciously attend to the act of mapping the sounds of spoken language to letters on the printed page in order to read words because that level of conscious, painstaking effort is no longer required to complete the task of reading a word, sentence, or perhaps even paragraph. It is also worth noting that the acquisition of automaticity—as it applies to reading or any other learned skill—is not comprised of discrete stages, nor is it dichotomous, but rather exists on a continuum so that one process (e.g., letter-sound correspondences) may be more automatic than another process (e.g., decoding) and less automatic than a third (e.g.,

reading connected text).

Relevance of Unitization and Automaticity to This Study

Students' ability to read words as whole units and to decode words with automaticity is relevant to this study because one of the outcome variables of interest, although not an original score in the DIBELS 6th edition revised version of Nonsense Word Fluency, is the number of words a student read correctly as a whole unit (i.e., phonologically recoded). Although this outcome variable has since been incorporated into the latest edition of the DIBELS (Good & Kaminski, 2011), and the scoring procedures are slightly different, little research has been conducted to examine the utility of this new score (Laugle, 2009). Second, examining the role of instruction on students' ability to decode nonsense words on a commonly used screening and progress monitoring measure may provide teachers with valuable information about student performance, as research has indicated that students who are able to read nonsense words as whole units typically perform better on later reading tasks, such as measures of oral reading fluency (Cummings, Dewey, Latimer, & Good, in press; Harn et al., 2008).

Examining School Effects

Although the quality of the reading instruction that students receive undoubtedly plays a critical role in the development of their early literacy skills, so too is it necessary to consider and examine the context within which that instruction is provided—schools. Perhaps one of the most notorious efforts made by researchers to examine the efficacy and effectiveness of schools for children was the research conducted by Coleman and his

colleagues (1966) to examine the availability of educational opportunities for students from different racial, ethnic, and socioeconomic backgrounds. Although the findings obtained from this study led Coleman et al. (1966) to claim that the funding allocated to schools had little effect on student achievement and that schools made no difference, there is now a widespread assumption that schools *do* affect children's development and that there are observable regularities in schools that add value to that development (D'Agostino, 2000; Reynolds, Teddlie, Creemers, Scheerens, & Townsend, 2000). D'Agostino (2000), for example, in his review of the literature located several studies indicating that school-level variables (e.g., socioeconomic status) have had direct effects on student achievement above and beyond effects associated with teachers and classrooms.

Comprehensive School Reform Models

Various waves of school reform have occurred during the last five decades, such as top-down reforms that have focused on systemic change, as well as those that have addressed the relationships between schools and families and promoted the needs of special groups of students (Desimone, 2002). These reforms, which have been characterized by the uncoordinated implementation of a series of specialized programs and initiatives, have been implemented in response to claims that schools have minimal positive impact on student learning (Borman, Hewes, Overman, & Brown, 2003; Desimone, 2002), and have resulted in minimal changes in school organization and the way teachers teach. Comprehensive School Reform (CSR), the most recent approach to schoolwide reform that has been supported by approximately $530 million from 1998-

53

2002, with an additional $21 million set aside for an independent evaluation of CSR programs, demonstrates the continued belief of educators and policymakers alike that schools matter (Desimone, 2002).

Unlike previous school reform efforts, implementation of CSR has been targeted in schools that have demonstrated the greatest need for reform and improvement (i.e., schools with high poverty rates and low student achievement scores) and instead of focusing on a number of specialized, uncoordinated initiatives focuses instead on reorganizing and improving *entire* schools (Borman et al., 2003; Desimone, 2002). CSR efforts, such as Title 1, the Elementary and Secondary Education Act of 1965 and its recent reauthorization as the No Child Left Behind Act of 2001 (P.L. 107-110) have, as discussed earlier in this chapter, received substantial amounts of financial support for their widespread implementation in schools in the United States. These schoolwide reform efforts, however, did not emerge as a prominent strategy for helping improve the outcomes of low-performing students in high-poverty schools until the 1990s (Borman et al., 2003). This type of comprehensive, scientifically based reform effort, critical to this study because participating schools implemented a schoolwide reform of early literacy instruction, has been described by the U.S. Department of Education as being comprised of the following 11 components (Borman et al., 2003; Comprehensive School Reform Quality Center [CSRQC], 2006a):

1. Employs proven methods for student learning, teaching, and school management that are founded on scientifically based research and effective practices that have been replicated successfully in schools.

2. Integrates instruction, assessment, classroom management, professional

development, school management, and parental involvement.

3. Provides continuous, high quality professional development for all school staff (e.g., teachers, instructional assistants, school leaders, etc.).

4. Includes operationalized, measurable goals for student achievement and establishes benchmarks for examining student progress toward those goals.

5. Provides support for teachers, principals, administrators, and other staff by creating shared leadership and responsibility for the implementation of reform efforts.

6. Is supported by teachers, principals, administrators, and other staff throughout the school.

7. Uses high-quality external and technical support and assistance from an entity with expertise and experience in schoolwide reform and improvement, such as an institute of higher education.

8. Identifies the available federal, state, local, and private financial and other resources that schools can use to coordinate services that help support and sustain the reform effort.

9. Includes an infrastructure that encourages and supports the meaningful involvement of parents and the local community in planning, implementing, and evaluating school improvement activities.

10. Includes a plan for annual evaluation of the implementation of the reform and student outcomes achieved.

11. Is supported by the findings of scientifically based research and/or strong evidence that the program will significantly improve the academic achievement of participating children.

Moreover, recent examinations of the effectiveness of CSR models for student performance have revealed that schools implementing CSR models can have a statistically significant and meaningful effect on student performance (Borman et al., 2003; CSRQC, 2006a, 2006b). Borman et al. (2003), for example, in their meta-analysis of 29 widely implemented CSR models, found three models with strong evidence of effectiveness (Direct Instruction, School Development Program, and Success for All with effect sizes of $d = .21$, $d = .15$, and $d = .18$, respectively), three models with highly promising evidence of effectiveness, and two models with promising evidence of effectiveness; effectiveness of the models was determined by examining the quality and quantity of evidence related to each model as well as the provision of positive, statistically significant achievement results for students. These findings led Borman et al. (2003) to conclude that the effects of CSR models, overall, are statistically significant, practically meaningful, and appear to be greater than other interventions with similar purposes and intended for similar populations of students. These findings were later supported by meta-analyses conducted by the Comprehensive School Reform Quality Center (CSRQ Center) of CSR programs implemented in elementary, middle and high schools. Studies of the 22 elementary CSR models reviewed, for example, revealed that nine models had moderate ($d = +0.15$ to $+0.19$) to moderately strong effects ($d = +0.20$ to $+0.24$) on student achievement (CSRQC, 2006b), while only five of the 14 models implemented in middle and high schools were found to have moderate positive effects on student achievement (CSRQC, 2006a).

Although the comprehensive reviews conducted by the CSRQ Center did not attempt to summarize or synthesize the characteristics and qualities attributable to

successful CSR programs, these efforts made by Borman et al. (2003) provided some initially surprising results. In particular, these researchers (Borman et al., 2003) found that ongoing professional development, the establishment of measurable goals and benchmarks for student learning, staff support for the model, and the use of specific curricular and instructional materials accounted for very little of the variance observed in student performance. In fact, the only clear statistically significant predictor of student performance was the active involvement of parents and the community in school governance. These findings are surprising in light of the fact that these insignificant predictors are four of the 11 components integral to a comprehensive approach to schoolwide reform. Borman et al. (2003) hypothesized that rather than taking these findings to mean that these components don't matter, their non-significance may be a byproduct of differences in the *implementation* of these components that warrants further investigation.

<div align="center">Multi-Tier, Schoolwide Prevention Models</div>

One widely implemented and commonly accepted model of schoolwide reform that has gained recent attention in light of legislation incorporating a Response to Intervention approach to reduce the incidence and prevalence of reading difficulties in young children has been a multi-tiered, prevention-oriented service delivery model for reading instruction (Clements & Kratchowill, 2008; Fien et al., 2009; Greenwood, Horner, & Kratchowill, 2008). This approach, which models the approach to service delivery utilized in community and public health settings (Greenwood et al., 2008), has been developed in response to a need for schoolwide programs that incorporate factors

<div align="center">57</div>

critical to early literacy instruction, namely (a) effective reading instruction for all students; (b) early identification of students at risk for later reading difficulties; (c) adoption and implementation of effective interventions designed and implemented to meet students' needs; (d) professional development aligned with those interventions; and (e) efficient and effective development of school resources (e.g., finances, personnel, time, etc.) to sustain the program (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007).

<div align="center">

Combining CSR and Multi-Tiered Instruction:

The Schoolwide Reading Model

</div>

Although most multi-tiered models of instruction share the general features described earlier, of particular interest to this study is the Schoolwide Reading Model (SWRM), which requires the establishment of organizational components that are the same across all schools as well as alterable variables that allow schools to customize reading instruction to suit their context and meet the needs of their students (Coyne, Kame'enui, & Simmons, 2004); this is the context within which instruction was delivered by schools participating in this study. During its inception and initial implementation, the SWRM was comprised of four essential components to promote the "science of reading instruction": (a) the establishment of long-term reading goals and interim performance benchmarks to monitor student progress toward those goals; (b) the establishment of a schoolwide assessment system that monitors student *and* school performance in beginning reading; (c) the coordination of differentiated instructional interventions to meet the needs of all learners; and (d) a coordinated schoolwide schedule that helps

maximize critical resources, such as staffing, materials, and instructional time (Coyne, Kame'enui, & Simmons, 2004). Over time, however, the conceptualization of the SWRM has expanded to include the following seven essential components (Baker, et al. in press):

1. The adoption of schoolwide priorities and implementation of practices that focus on the content essential to reading development (i.e., phonological awareness, alphabetic understanding, fluency with connected text, vocabulary, and reading comprehension).

2. The periodic collection of reliable and valid assessment data to inform instructional practices (the frequency of data collection is determined by students' level of risk for later reading difficulties).

3. The establishment of a schoolwide schedule that allocates and protects sufficient time for reading instruction to ensure that students reach key reading goals.

4. An emphasis for all staff on high-quality implementation of research-based instructional programs.

5. The provision of differentiated, multi-tiered instruction designed to meet the needs of individual students.

6. The use of student performance data by school-level leadership to support effective classroom instruction with a focus on sustained, effective implementation.

7. The provision of high-quality professional development to continuously drive the schools' ongoing efforts to improve the quality of reading instruction and student achievement.

Although these are the basic tenets of the SWRM, the model was slightly modified to adhere to the stringent guidelines set forth by Reading First legislation. These

modifications to the model included (a) a minimum of 90 minutes of literacy instruction

provided daily that was protected from interruptions (i.e., assemblies, field trips, etc.) in

the school schedule; (b) the use of comprehensive reading measures (the SAT-10 in

Grades K-2 and the Oregon Assessment of Knowledge and Skills [OAKS] in Grade 3) to

evaluate student progress at the end of each grade; and (c) a minimum of 30 minutes of

teacher-directed, small-group instruction provided daily for all students (Baker et al., in

press).

Apart from the modifications to the SWRM described above, each of these

components is a derivation of those included in the definition given by the U.S.

Department of Education of a scientifically based approach to Comprehensive School

Reform (Borman et al., 2003). Implementation of the SWRM, for example, requires the

adoption and implementation of scientifically based comprehensive core, supplemental,

and intervention programs to provide differentiated instruction to meet the needs of all

learners. While a comprehensive core program is a set of systematically designed and

sequentially aligned teacher and student instructional materials that focus on the five

critical components of early literacy instruction identified by the National Reading Panel

(NICHD, 2000), supplemental programs are designed to provide explicit instruction and

numerous opportunities for practice in one or two of the critical components of reading

(LLSSC & ODE, 2009). Intervention programs may vary in the scope of the content

covered (i.e., all five critical components or maybe just one or two) and are typically

characterized by any or all of the following features: (a) greater amounts of instructional

time; (b) explicit and systematic instructional design; (c) a focus on mastery learning (i.e.,

ensuring that students perform to a certain criteria before introducing new content); (d)

careful and frequent monitoring of progress toward formative reading goals; and (e) frequent delivery of instruction to small groups of students to afford them more opportunities to respond and receive corrective feedback (LLSSC & ODE, 2009). The adoption and implementation of these programs using explicit, systematic teaching practices aligns with the first component of CSR programs (employment of proven methods based on scientifically based research).

In addition, teachers and other instructional staff use these materials to provide students with the instruction needed to achieve grade-level reading goals and the quarterly benchmark goals established by the formative assessment systems (e.g., DIBELS) used to monitor progress toward those goals. The reading measures used as grade-level outcome measures and periodic benchmarking tools to monitor student progress, moreover, are required to be technically adequate (i.e., reliable and valid) to provide consistent indicators of student performance that enable teachers and leaders to make appropriate and trustworthy inferences about student performance. Furthermore, similar to CSR models, the SWRM includes the periodic provision of high-quality, research-based professional development for teachers and school leaders alike that is focused on critical SWRM components, such as the implementation of scientifically based research programs and instructional strategies with fidelity (i.e., as they were intended by program designers), the collection of various sources of student performance data to enable informed instructional decision-making, and the use of student performance data to modify instruction to meet the needs of all students.

Within the context of Oregon Reading First, all of these efforts were supported by frequent and intensive internal and external technical assistance and support, also a

61

component of CSR models. Internal support was provided by highly trained literacy coaches, and external support was provided by monthly (at a minimum) visits from regional coordinators extensively trained in implementation of the SWRM. In addition to providing technical support, the Oregon Reading First Center (ORFC) was also charged with evaluating implementation of the SWRM not only during monthly schools but also via semiannual examination of student reading outcomes and the dissemination of information about student performance to schools. Purposeful comparisons of the SWRM and CSR, in other words, reveal extensive similarities between the two models; six of the 11 components are similar at the school level.

It is also worth noting that other critical components of a CSR program that are beyond the scope of the school's implementation are also a part of the SWRM, such as the establishment of a relationship with a knowledgeable and experienced external organization to obtain high-quality support and technical assistance and the identification of funding sources (in this case, federal) to support the coordination of services and implementation of the model. Additionally, all staff in schools (including principals and district leadership) participating in this study (and in the Oregon Reading First initiative in general) were required to sign letters of commitment annually to demonstrate their support for and agreement to implement the model, and schools were supported in the establishment of a culture of shared leadership through the hiring of a literacy coach and through efforts guided by the technical assistance center (in this case, the ORFC) to convene and meet periodically with grade-level teams of teachers and school-level early reading teams.

Examining the Effectiveness of the Schoolwide Reading Model in Changing Student Outcomes

Although the SWRM is grounded in research-based components (Coyne, Kame'enui, & Simmons, 2004) and therefore could be expected to have a positive effect on student outcomes, empirical support for the model is available via positive outcomes observed in student achievement for schools who have adopted this schoolwide, data-driven, prevention-oriented approach to reading instruction (Baker et al., in press; Chard & Harn, 2008; Chard et al., 2008; Fien et al., 2009). Baker et al. (in press), for example, examined whether school experience with implementation of SWRM within the context of Reading First was associated with positive student outcomes by comparing the student outcomes of Oregon Reading First schools that had implemented the model for 2 complete years (Cohort A; $N = 34$) with those of Oregon Reading First schools that had implemented the model for only 1 year (Cohort B; $N = 17$). Schools participating in both cohorts met the same eligibility criteria, received the same extensive professional development focused on implementation of the SWRM, and began their participation in the initiative with similar ranges of students requiring explicit and systematic instructional support, the only difference being the year that their participation in the federally funded initiative began (2003-2004 for Cohort A schools, 2005-2006 for Cohort B schools; Baker et al., in press).

Despite the fact that specific numbers are not provided, examination of student performance on early indicators of literacy (DIBELS) and comprehensive measures of literacy (SAT-10 and OAKS) revealed generally that students improved over time in both cohorts of schools and while students in the Cohort A schools outperformed students in

the Cohort B schools during the year of interest (2005-2006), the gaps observed in student performance were beginning to decrease by 2007-2008 (Baker et al., in press). Furthermore, comparisons of performance across the two cohorts of schools revealed that after 2 years of implementation, Cohort A schools had outperformed Cohort B schools on early indicators (as evidenced by Cohen's d effect sizes of +0.28 for NWF in the winter of kindergarten, +0.26 for ORF in the fall of first grade, +0.20 for ORF in the fall of second grade, and +0.23 for ORF in the fall of third grade) and comprehensive measures (+0.34 for the SAT-10 in kindergarten, +0.18 for the SAT-10 in first grade, +0.19 for the SAT-10 in second grade, and +0.31 for the OAKS in third grade) of reading. Taken together, these findings support the effectiveness of the SWRM in increasing student outcomes in reading and demonstrate that schools may have varying effects on student achievement depending on the amount of time they have been engaged in a schoolwide reform effort.

Also of interest is a study conducted by Fien et al. (2009), who were interested in whether schools implementing the SWRM had positive effects on student achievement and in the amount of variation in early reading outcomes for students in kindergarten (DIBELS PSF and NWF) that existed between schools. These research questions were examined using student and school-level data obtained from 57 elementary schools in Hawaii that received training in and provided instruction within the context of the SWRM. As noted earlier, the student outcomes of interest were DIBELS PSF and NWF in kindergarten, and the school-level context variables that the researchers considered were the average socioeconomic status of schools and their scores on a measure designed to reflect the degree to which school staff believed there were policies and practices in

64

place to support effective beginning reading instruction (Planning and Evaluation Tool-Revised [PET-R]; Kame'enui & Simmons, 2003). Fien et al. (2009) found that some schools were able to get 100% of students to meet the PSF benchmark goal and 100% to meet the progressive benchmark goal for NWF, indicating that schools implementing the SWRM did have positive effects on student achievement.

Analyses of the data also revealed that in the unconditional model with no student and school-level predictors, 22% and 36% of the variance observed in PSF and NWF outcomes, respectively, existed (Fien et al., 2009). When school-level predictors (percentage of students categorized as low risk in the winter and the pretest score obtained on the PET-R) were included in the model, 40% of the variance observed between schools in students' PSF and NWF scores was explained, providing Fien et al. (2009) with empirical support that substantial differences between schools with similar contexts existed and that reliable and meaningful differences in student outcomes could be observed between those schools. The findings of this study, in other words, not only provide evidence that implementation of a schoolwide, multi-tiered, prevention-oriented approach to early reading instruction can produce positive outcomes for students, but also that relatively large proportions of the variance observed in student performance can be explained by the schools within which students are enrolled, lending support to the idea that schools *do* matter and are a worthwhile unit of study.

The Importance of Different Types of School Effects

Although it is undoubtedly reassuring to see that schools can and do have positive effects on student outcomes, when examining those effects it is also necessary to be clear about which types of effects are being considered, as multiple definitions of school effects and effectiveness exist. Teddlie et al. (2000), for example, put forth six different definitions of school effects, each of which is related to different methodological issues (e.g., reliability, validity) in educational research. Of particular interest to this study is the understanding of school effects as measuring the extent of variation *between* schools in the total variation observed in individual students' test scores (Teddlie et al., 2000). The purpose of these types of studies is to examine what percentage of total variation in students' scores on some measure of academic achievement (in this case, DIBELS NWF as an indicator of alphabetic understanding) is "between schools" as opposed to some other level in the mathematical model; the researchers report that findings from these studies typically indicate that between 8% and 15% of the variance observed in students' test scores can be attributed to school-level variables (Teddlie et al., 2000). Within the context of the SWRM and CSR models, for example, one could investigate what percentage of total variation in students' scores could be attributable to schools implementing those models.

Type A and Type B Effects

A second conceptualization of school effects proposed by Raudenbush and Willms (1995) is also applicable to this study, as it breaks down the definition of a

"school effect" as the extent to which attending a particular school influences or modifies student outcomes into two types of school effects. Both conceptualizations of school effects are concerned with the difference between the performance of students in a particular setting compared to their expected performance if they had been in some other setting. According to Raudenbush and Willms (1995), a Type A effect represents the difference in students' observed performance on the measures of interest and their expected or predicted performance if they had attended a "typical" school. For this study, a Type A effect will represent the difference in students' observed performance on DIBELS NWF and their expected performance if they had attended a "typical" school that had not established a comprehensive, multi-tier system for reading instruction.

Type B effects, in contrast, expand this examination a bit further by taking into consideration school practice and school context variables. School practice is comprised of such variables as utilization of resources, curricular content, and administrative leadership, while school context is considered to include any variables that are exogenous to the practices of the school staff, such as the social and economic characteristics of the community in which the school is located (Raudenbush & Willms, 1995). Examination of Type B effects will be considered in this study via the inclusion of two school-level demographic variables—percentage of incoming kindergarten students at risk and the percentage of students eligible for free and reduced-price lunch—as indicators of school context. Although one might expect that schools from a relatively homogeneous population of schools (such as those participating in this study) might not account for substantial proportions of observed variance in students' performance because they are so similar, one objective of this study is to investigate this issue further, as prior research has

indicated that relatively homogeneous schools can account for significant amounts of variation in student performance (Fien et al., 2009).

Examining Student Characteristics

Although considering the role of instruction and the context within which it is delivered in students' acquisition and demonstration of alphabetic understanding is undeniably important as research has found that the nature and quality of classroom instruction a child receives can be a "pivotal force" in the prevention of reading difficulties (Felton, 1993; National Research Council, 1998), the skills and personal experiences a student brings to the classroom are also important. Of particular relevance to this study are the following student-level characteristics: (a) student's initial level of skill on DIBELS PSF & NWF, and (b) student's English language proficiency.

Initial Status on NWF

Just as the work of numerous researchers has confirmed the importance of the skills children bring with them to their journey of becoming skilled readers (Adams, 1990; National Research Council, 1998), so too have researchers concluded that the initial skill level of students is one of the strongest determinants of the growth they will make during the school year (Good, Simmons, & Smith, 1998; Hedges & Hedberg, 2007). Recently, several studies have been conducted to examine the role of students' initial skill level on the measure of interest in this study, DIBELS NWF (Fien et al., 2008; Fien et al., 2010; Good et al., 2009; Harn et al., 2008), in predicting variance on later reading measures (e.g., NWF, ORF). Fien et al. (2010), for example, examined the

role of initial skill status on NWF with a sample of approximately 3,400 first-grade

students enrolled in 50 schools participating in the federally funded Oregon Reading First

initiative during the 2006-2007 school year. Rather than using students' raw scores as an

indicator of their initial levels of performance on NWF, the researchers created five strata

using those raw scores that were based on the groups of scores that have been used to

determine different instructional recommendations (e.g., 0-12 CLS in stratum 1, 13-23

CLS in stratum 2, 24-50 CLS in stratum 3; 51-70 CLS in stratum 4; and 71 and above in

stratum 5). They found that approximately 48% of the variance in students' ORF scores

in the spring of first grade was accounted for by their initial skill status on NWF in the

fall of first grade (Fien et al., 2010). Fien et al. (2008) reported similarly high proportions

of variance explained by initial risk category and fall of first-grade NWF score in their

examination of data obtained from Oregon Reading First schools from 2003-2006. In this

instance they found that these variables accounted for approximately 58% of the variance

observed in students' ORF scores at the end of first grade.

Similar findings have also been obtained outside of Oregon Reading First schools.

Good et al. (2009), for example, conducted their study with two samples of students—

2,172 first grade students enrolled in Oregon Reading First schools during the 2004-2005

school year, and 358,032 first graders from 44 states in the United States and Canada

who had a complete set of NWF data (e.g., fall, winter, and spring) in the DIBELS Data

System (DDS) during the 2004-2005 school year—to examine their hypotheses about the

role of initial level of performance on NWF in partitioning variance on a measure of

accuracy and fluency with connected text (i.e., DIBELS Oral Reading Fluency). Using

multiple regression analyses, Good et al. (2009) found that initial NWF score (i.e., fall)

accounted for 58% of the variance for students enrolled in Oregon Reading First schools and 50% of the variance for students whose scores were obtained from the DDS.

<center>Is Performance on NWF Different for English Learners?</center>

Due to the rapidly growing population of English Learners (ELs) receiving educational services in schools throughout the United States, and Oregon in particular, it also makes sense to examine whether ELs perform differently on NWF compared to their native English-speaking peers. Figures released during the National Symposium on Learning Disabilities in English Language Learners (USDOE & NICHD, 2003), for example, indicated that since 1980 the number of Hispanic children enrolled in America's schools has increased from 9% to 16%, and the number of Asian and Pacific Islander children has doubled, from 2% to 4%. Others (Fitzgerald, 1993; Gunn, Smolkowski, Biglan, Black, & Blair, 2005) have reported that ELs currently account for approximately 6% of the school-age population and estimated that Spanish-speaking children in particular represent between 70% and 80% of that population. The U.S. Department of Education and National Institute of Child Health and Human Development (USDOE & NICHD, 2003) also estimate that approximately 40% of all students enrolled in the nation's schools will be ELs by the year 2030. Moreover, 77,216 of Oregon's 552,505 students, or 14%, are categorized as Hispanic or Latino, and of those 77,216 students, 53,364, or 9.5%, were identified as ELs (NCES, 2010c). This represents an increase of 106.9% increase in the number of Hispanic/Latino students attending Oregon schools from 1998-1999 to 2008-2009.

Research has indicated, however, that in the early years of school English learners

(ELs) can learn to read in English as well as their native English-speaking peers (August & Shanahan, 2006; Gersten, Baker, Haager, & Graves, 2005; Lesaux & Siegel, 2003). Gunn et al. (2005) demonstrated this fact in their comparison of supplemental versus no supplemental reading instruction for Hispanic and non-Hispanic students in Grades K-3. During the first 2 years of this study, students who were randomly assigned to supplementary reading instruction groups received approximately 15 to 16 months of intensive, explicit, and systematic instruction that focused on the development of fluent word recognition skills via instruction in phonemic awareness, phonics, and practice reading decodable text. The effects of the intervention were examined on measures of letter-word identification, word attack (pseudo-word reading), and passage comprehension and results revealed the following: (a) Students receiving the intervention outperformed and exhibited greater rates of growth on letter-word identification and word-attack than nonintervention students at the end of the intervention, but these differences had faded by the end of the second year following intervention delivery; and (b) on measures of passage comprehension, intervention students scored higher than nonintervention students 1 year after participating in the intervention, but were performing comparably by the end of the second year following intervention delivery. These findings led the researchers to conclude that Hispanic students were able to benefit from the supplemental reading instruction as much as, if not more than, their non-Hispanic peers.

In addition to the wealth of research conducted recently examining the role of initial skill status on NWF in partitioning the variance in student's scores on later reading measures, research has also been conducted recently to examine if DIBELS NWF

functions similarly for English Learners (ELs) and native English speakers (Fien et al., 2008; Vanderwood, Linklater, & Healy, 2008). Fien et al. (2008), for example, investigated the validity of DIBELS NWF as an indicator of early reading proficiency for native English speakers and English learners with a large sample of students participating in the Oregon Reading First project from 2003-2006. To answer their questions the researchers correlated students' NWF scores with later Oral Reading Fluency (ORF) and SAT-10 reading comprehension scores, examined the correlations of ELs and native English speakers for any statistically significant differences, and correlated NWF scores across time and tested the correlations for the two groups for differences in stability. These analyses revealed that, for all students, performance on NWF accounted for between 26% and 58% of the variance in ORF and between 31% and 54% of the variance in SAT-10 scores. Examination of differences between the correlations for ELs and native English speakers revealed that only five of 24 were statistically significant, thereby indicating that NWF appears to function similarly for both student groups and that the correlations remained similarly stable for both groups over time (Fien et al., 2008).

In addition, researchers have recently examined the contribution of English Learners' (ELs') initial skill status and growth on Spanish word-word reading in kindergarten and first grade on later measures of Spanish reading comprehension. Although not directly related, this line of research is still relevant not only because of its focus on ELs but also because it provides information about the degree of influence the skills and experience a student brings to instruction will have on their reading skills measured in their native language. Baker, Park, and Baker (2010) conducted a study with 168 kindergarten and first-grade students, all of whom spoke Spanish as their primary

language at home and were classified as Limited English Proficient by Oregon state standards. These students, in addition to receiving English literacy instruction, received a minimum of 90 minutes of Spanish reading instruction; 30 minutes of this instruction was provided in homogeneous small groups.

Results indicated that although students' initial skill status on a measure of Spanish pseudo-word reading, FPS, did not have a significant, direct prediction of students' reading comprehension skills at the end of first grade, it did indirectly predict first-grade reading comprehension via students' performance at the end of kindergarten. That is to say, a direct relationship was observed between students' initial status on FPS in the winter of kindergarten and their reading comprehension scores at the end of kindergarten, and their later performance on measures of reading comprehension at the end of first grade; this model, in fact, explained approximately 53% of the variance in students' reading comprehension scores at the end of Grade 1. Although these results aren't as directly conclusive as those from previously discussed studies in regard to the contributions made by students' initial skill status, the findings nonetheless indicate that the skills students do bring with them to their educational experience are important (Adams, 1990; National Research Council, 1998; NICHD, 2000).

Although several studies cited here examined the performance of ELs on measures of Nonsense Word Fluency from different perspectives and additional research has indicated that ELs can acquire English literacy skills at the same rate of their native English-speaking peers (Gersten et al., 2005; Gunn et al., 2005), examining the importance of this student-level effect within the context of this study is still relevant because of the continually increasing numbers of ELs receiving educational services in

our elementary schools today (USDOE & NICHD, 2003).

Examining Measurement Properties

Examining the efficacy and effectiveness of the instructional supports provided, however, as well as the effect of the context of instruction (i.e., school) on a student's demonstration of any skill also depends on the use of technically adequate measures to evaluate student progress. According to the *Standards for Educational and Psychological Testing* (*Test Standards;* AERA et al., 1999), test developers have the responsibility to develop tests that permit valid interpretations of student performance about the construct of interest and to ensure that those tests can be used to obtain consistent results about student performance on the construct of interest over time and/or across different populations of students. Each of these elements of technical adequacy—validity and reliability—can be supported by the collection of different types of evidence, the following of which are of particular interest to this study: (a) construct validity; (b) sources of convergent evidence (i.e., predictive and concurrent validity); and (c) indices of reliability, or consistency of measurement (i.e., alternate form and test-retest reliability). A review of the literature on measures designed to assess student knowledge of the alphabetic principle is laden with the first four sources of evidence that support the use of pseudo-word reading measures to examine students' acquisition of alphabetic understanding.

Construct Validity

According to Cronbach and Meehl (1955), a construct can be defined as "some

74

postulated attribute of people, assumed to be reflected in test performance" (p. 283). Test developers need to engage in the process of construct validation during the creation of all tests and measures of psychological constructs because they are not directly observable but are *assumed* to be accurately and appropriately represented by those tests and measures. To examine whether a construct has been appropriately represented on a test or measure, Cronbach and Meehl (1955) recommend utilizing any of the following methods: (a) comparison of group differences, (b) correlations between measures presumed to measure the same construct, (c) examination of the stability of test scores over different testing occasions, and (d) examination of the percentage of variance attributable to the construct variable. A review of the literature reveals that several sources of evidence are available to indicate that there is consensus in the field that measures of pseudo-word reading are appropriate for examining alphabetic understanding, and that the measures available are stable and provide consistent indicators of student performance over time.

How Best to Measure Demonstration of Alphabetic Understanding?

Researchers have observed, for example, that the measurement of this construct is challenging because there are two components of the alphabetic principle—the cognitive processes associated with learning and understanding the relations between letters and words, and the *application* of that understanding to read words—neither of which can be measured directly (Laugle, 2009). Despite this challenge, a review of the literature indicates that the most common measure of students' alphabetic understanding, for native English-speaking and English learners alike, are measures of pseudo-word reading (Chard et al., 1998; Lesaux, Koda, Siegel, & Shanahan, 2002; NICHD, 2000; Siegel,

1998; Vellutino, 1991). Siegel defined pseudo-words as words composed of a pronounceable combination of letters that can be read via the application of grapheme-phoneme correspondence rules but are not real words in English; Juel (1991) provided an even simpler definition that makes an even more important distinction: pseudo-words are pronounceable and irregular, while nonwords are not.

Examination of the meta-analyses conducted by the National Reading Panel (NICHD, 2000), for example, revealed that of the 38 studies included that examine the effectiveness of phonics interventions, 18 (or 47%) used a measure of pseudo-word reading to determine the impact of the intervention. The most frequently cited reason for using measures of pseudo-word reading for examining students' acquisition of the alphabetic principle is that these tests specifically isolate students' ability to apply letter-sound correspondences and blending from word meaning (Chard et al., 1998). It has also been argued that an advantage of pseudo-word reading measures is that they avoid tapping the other word reading skills (e.g., memorization, analogy, etc.) students may use to read real words by forcing them to apply and demonstrate their knowledge of letter-sound correspondences and phonological recoding (Fien et al., 2008, Good et al., 2009). Furthermore, students' performance on measures of pseudo-word reading is one of the best predictors of word identification and reading ability (Curtis, 1990; Vellutino, 1991). The findings from these studies, in other words, provide substantial support for the use of measures of pseudo-word reading, such as DIBELS NWF, as a measure of alphabetic understanding.

Sources of Convergent Evidence: Concurrent and Predictive Validity

A second source of potential construct validity proposed by Cronbach and Meehl (1955) are the correlations that can be predicted between two measures that are hypothesized to measure the same construct. More specifically, the *Test Standards* (AERA et al., 1999) describe one source of validity evidence that can be used to support one's argument that a measure appropriately represents the construct of interest: convergent validity. Evidence of convergent validity, or the relation between test scores and other measures designed to assess similar constructs, can be supported by indications of predictive and/or concurrent validity. Predictive validity refers to how accurately performance on the test of interest predicts later performance on a measure targeting a similar construct at a later point in time, whereas concurrent validity refers to how accurately performance on the test of interest predicts performance on a measure targeting a similar construct at about the same time (AERA et al., 1999). Each of these sources of evidence supports the construct validity of the test in question by demonstrating that students' performance on the test in question is similar to their performance on other similar measures; if the tests were not measuring the same construct, one would not expect students' to perform similarly or obtain similar scores.

Examination of a recently released technical adequacy supplement for all DIBELS 6th edition measures (DMG, 2008) summarizes the results of numerous studies conducted in the last decade that demonstrate moderate to strong evidence of convergent validity for DIBELS NWF. Researchers have obtained, for example, concurrent validity coefficients ranging from .68 to .75 between DIBELS NWF and the Test of Word

77

Reading Efficiency (TOWRE) Phonemic Decoding Efficiency, a timed measure of pseudo-word reading in which pseudo-words increase in their difficulty and length (including multisyllabic pseudo-words; Burke & Hagan-Burke, 2007; Hagan-Burke, Burke, & Crowder, 2006). Although no predictive validity coefficients for DIBELS NWF and other measures of pseudo-word reading were reported, predictive validity coefficients with other indicators of reading performance (e.g., Oral Reading Fluency) and comprehensive measures of reading (e.g., Iowa Test of Basic Skills, Stanford Achievement Test, etc.) of low (.27) to moderate (.71) were reported, providing evidence that students' performance on DIBELS NWF can be somewhat predictive of their performance on later measures of reading (DMG, 2008).

Indicators of Measurement Consistency: Alternate Form and Test-Retest Reliability

A third recommendation put forth by Cronbach and Meehl (1955) for investigating construct validity is to examine the stability of test scores over time. Also referred to as reliability, the *Test Standards* (AERA et al., 1999) propose two methods for examining the consistency of a test when it is used with the same population of individuals or groups: (a) alternate form reliability, consistency in scores across parallel forms of a test; and (b) test-retest reliability, consistency in scores across different measurement occasions using the same form of a test. Examination of these and other sources of reliability evidence are needed to ensure that the amount of measurement error associated with a particular test (or test form) is sufficiently low enough that the generalizability of the test results (or students' performance) is not affected (AERA et al., 1999). Recent studies, for example, have produced alternate-form reliability coefficients

ranging from .67 to .94 (DMG, 2008), indicating that the scores students obtain on alternate forms of NWF have demonstrated moderate to strong stability, thereby supporting inferences teachers can make about students' knowledge and skills.

Examination of Item Variability

Although theoretical and empirical evidence is available to support the use of total raw scores (Correct Letter Sounds and Whole Words Recoded Completely and Correctly) to make appropriate inferences about students' knowledge of the alphabetic principle, a review of the literature revealed that minimal research has been conducted examining the variability of item difficulty for any DIBELS measures (e.g., Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, etc.).

To date, much of the research conducted examining the possibility of measurement variance associated with indicators of early literacy, such as DIBELS or other curriculum-based measures, has focused on the invariability of ORF passages within a grade (Ardoin & Christ, 2009; Betts, Pickart, & Heistad, 2009; Francis et al., 2008; Petscher & Kim, 2010; Poncy, Skinner, & Axtell, 2005). Within this context, in which the primary interest is appropriately modeling students' growth, measurement invariance is necessary to obtain accurate scores of students' true reading rates because biased estimates of performance are likely to produce inaccurate representations of students' growth (Francis et al., 2008).

Francis et al. (2008), for example, investigated the comparability of items (i.e., passages) by having 134 second-grade students read six randomly selected and randomly ordered DIBELS ORF progress monitoring passages; students read three passages during

79

the first administration and the remaining three passages over an additional 6 weeks. Results revealed a lack of passage equivalency across the passages included in the study, as evidenced not only by the range of fluency rates (and their associated standard deviations) for each passage, but also by the significantly different error variances associated with each passage. Petscher and Kim (2010) also found that anywhere from 2% to 9% of the variance observed in first- through third-grade students' ORF passages could be attributed to potential passage effects, or a lack of measurement invariance associated with the ORF passages.

This information regarding passage equivalency is valuable for two distinct groups of educators: (a) researchers, methodologists, and test developers, and (b) classroom teachers and school-level leaders. This information is valuable from a measurement perspective because it provides researchers, methodologists, and test developers with information needed to determine how best to equate passages and develop passage sets of comparable difficulty. Moreover, if the passages are of comparable difficulty (i.e., some passages are not significantly easier or difficult than others), the scores obtained from these passages will permit more valid and appropriate inferences about student performance. In addition, this information is valuable to classroom teachers and school-level leaders because it may provide some insight as to why students may not obtain consistent scores across passage sets during different test administrations. Although this notion of passage equivalency and measurement invariance in ORF passages has received increased attention in recent years (Ardoin & Christ, 2009; Betts et al., 2009; Francis et al., 2008; Petscher & Kim, 2010; Poncy et al., 2005), a review of the literature revealed no similar studies examining variability of item

difficulties of other early literacy indicators such as DIBELS NWF, although examination

of variability in item difficulty may also have important measurement and instructional

utility as well.

CHAPTER III

METHODS

The purpose of this study was to examine school, student, and measurement
effects on 1,111 students' demonstration of the alphabetic principle via their performance
on the DIBELS Nonsense Word Fluency (NWF) measure in the spring of first grade.
These first-grade students were enrolled in 14 elementary schools participating in the
federally funded Oregon Reading First initiative during the 2008-2009 and 2009-2010
school years.

More specifically, this study was guided by the following four objectives: (a) to
test the hypothesis that school context variables (such as academic achievement and
socioeconomic status) account for variation in two outcome scores associated with NWF,
(b) to model between- and within-school variation based on these context variables, (c) to
examine the degree to which student-specific variables account for variation in these
same NWF scores, and (d) to examine whether the difficulties of the pseudo-words on the
spring of Grade 1 NWF benchmark probe vary significantly. The purpose of this study
was to answer the following research questions:

1. How much variance in students' performance on DIBELS NWF at both the
sound (total Correct Letter Sounds) and word (Words Read as Whole Units Correctly)
levels exists between and within schools? How much of that variance is attributable to the
following school context variables: (a) the poverty level of the school, and (b) the
percentage of incoming kindergarten students categorized as being at risk for later
reading difficulties?

2. How much of the variance in student performance on DIBELS NWF at both the sound and word levels is accounted for by student-level characteristics, such as student EL status and initial performance on measures of Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF)?

3. Is there significant variation in the item difficulty of the pseudo-words on the spring of Grade 1 NWF benchmark probe? If so, do these items vary significantly as a function of the following possible sources of variance: student-level characteristics (English language status and initial scores on Phoneme Segmentation Fluency, and Nonsense Word Fluency)?

Participants

As noted earlier, participants in this study were first-grade students attending 14 elementary schools that participated in the federally funded Oregon Reading First initiative during the 2008-2009 school year and first-grade students attending five participating elementary schools during the 2009-2010 school year. Schools eligible to receive Reading First funding in Oregon were required to meet specific criteria for student poverty level and performance on measures of reading. For the 2008-2009 school year, between 55.9% and 92.5% of students in each participating school were eligible for free and reduced-price lunch services (a common index of socioeconomic status); data for the 2009-2010 school year are currently unavailable. Additionally, the percentage of incoming kindergarten students categorized at risk for later reading difficulties based on their performance on DIBELS measures administered in the beginning of kindergarten

83

(an index of school risk status used during Oregon Reading First to make comparisons across schools of similar populations) ranged from 15% to 64%.

To be eligible for participation in this study, students had to have the following data: (a) a score on PSF from the fall of Grade 1, (b) a score on NWF from the fall of Grade 1, (c) indication in the DIBELS Data System (DDS) of whether they were an English Learner, and (d) *both* outcomes of interest on NWF (CLS and WRWUC) in the spring of Grade 1. To this end, 87 of the originally eligible students were removed from the sample due to a lack of information about their EL status, and one influential case with a fall NWF score of 213 was removed in order to obtain the best estimates possible about the relations of interest. Of the 951 students eligible for inclusion from the 2008-2009 school year, approximately 44% were female, 43% were Caucasian (although ethnicity data were missing for approximately 100 students), and 22% were classified as ELs (again, data were missing for approximately 90 students). Of the 365 students eligible for inclusion from the 2009-2010 school year, approximately 41% were female, 37% were Caucasian, and 28% were classified as ELs.

Although 14 schools participated in the Oregon Reading First initiative during the 2009-2010 school year and it is likely that a sizeable proportion of those students have both PSF and NWF scores from the fall of Grade 1, there is one reason why students from these schools weren't eligible for participation in this study: student DIBELS NWF data had to be collected on specialized forms (see Appendix) that allowed for explicit calculation of the WRWUC variable. Additionally, it is worth noting that for this study, rather than conducting separate analyses for each of the two participating cohorts, I combined the data from the two years into one data file for ease of interpretation.

Although this means that the number of participating students in five of the 14 schools was larger than the others, this decision seemed appropriate because (a) minimal changes were observed between school years in the school context variables; and (b) minimal changes were evident in the support received by schools to implement the research-based, multi-tiered reading instruction model adopted by all participating schools.

Commonalities Across Oregon Reading First Schools

To be eligible for participation in the Oregon Reading First initiative and receive extensive technical assistance from the Oregon Reading First Center, districts were required to meet certain poverty and student performance criteria. Districts had to meet one of the three following conditions in order to meet the poverty criterion: (a) The district needed to qualify for Title I School Improvement; (b) the district needed to qualify as an Entitlement Empowerment Zone by the U.S. Department of Agriculture (Billings, 2002); or (c) at least 20% of students, or more than 1,000 students total, had to be from families with incomes below the poverty line (Billings, 2002). The student performance criterion required that more than 21% of third-grade students, or more than 100 third-grade students in the district total, scored below the state standard for reading proficiency as measured by the reading subtests of the OAKS (Billings, 2002). Only schools within districts that met these criteria were eligible to participate in Oregon Reading First and had to meet a separate set of poverty and student performance criteria: (a) Either the school had to qualify for Title I School Improvement or a minimum of 50% of all students had to qualify for free and reduced-price lunch services according to federal guidelines; and (b) more than 21% of third-grade students (or more than 15

students total) scored below the proficiency standard on the state reading assessment during the previous school year (Billings, 2002).

Additionally, Reading First schools in Oregon established and implemented the Schoolwide Reading Model (SWRM) as a school-level approach to providing differentiated reading instruction based on scientific research in reading. This model had not only been widely implemented and refined prior to Reading First (Coyne, Kame'enui, & Simmons, 2004; Simmons, Kuykendall, King, Cornachione, & Kame'enui, 2000), demonstrating its feasibility and success, but was also closely aligned with the intended implementation of Reading First requirements (Baker et al., in press). The Schoolwide Reading Model as implemented by Oregon Reading First schools contained the following seven essential elements (Kame'enui, Simmons, & Coyne, 2000; Simmons et al., 2002):

1. Schoolwide priorities, practices, and instruction focused on the five essential components of beginning reading: phonological awareness, alphabetic understanding, reading fluency with connected text, vocabulary, and comprehension (NICHD, 2000).

2. Reliable and valid assessment data collected at least three times per year (e.g., fall, winter, and spring) and analyzed to inform instructional practices.

3. A minimum of 90 minutes of uninterrupted, protected time for whole group reading instruction plus an additional 30 minutes of small-group, differentiated reading instruction to make sure students met key reading goals and grade-level benchmarks.

4. Adoption and implementation of high-quality, scientifically based reading programs focused on the five essential components of beginning reading.

5. Differentiated, multi-tiered instruction designed to provide varying levels of support to meet students' individual needs.

86

6. Use of various sources of student performance data (e.g., DIBELS, in-program assessments, phonics screener, etc.) to support effective classroom instruction with a focus on sustained, effective implementation of instructional practices.

7. High-quality professional development is provided for all staff that is driven by student and staff needs and drives ongoing efforts to improve the quality of reading instruction and student achievement.

Although schools participating in the Oregon Reading First project varied in their implementation of these components of the Schoolwide Reading Model (e.g., reading programs adopted to provide instruction on the five essential components, length and structure of the uninterrupted reading block, decision rules used for placement and movement and students between the tiers of instruction, etc.), each participating school was expected to implement the aforementioned components to meet the requirements of Reading First with the goal of improving reading outcomes for all students in Grades K-3 (Baker et al., 2007).

Measures

To investigate the research questions of interest, two student performance measures were included in this study: DIBELS Phoneme Segmentation Fluency (PSF) and DIBELS Nonsense Word Fluency (NWF). Information about the administration and scoring procedures for each of these measures, as well as technical adequacy information, are described below. In addition, information about the modified scoring procedure used for NWF to obtain a Words Read as Whole Units Correctly (WRWUC) score for each student is also provided.

DIBELS Phoneme Segmentation Fluency (PSF)

DIBELS PSF (6th edition revised; Good & Kaminski, 2002) is a 1-minute, standardized, individually administered measure designed to assess students' phonological awareness by measuring their ability to segment three- and four-phoneme words into their individual sounds. The words are presented orally to students whose scores are recorded as the number of phonemes they provide for each word. If, for example, the word is "trick," the student could earn four points for producing all correct phonemes (e.g., /t/ /r/ /i/ /k/), three points for blending some of the phonemes (e.g., /tr/ /i/ /k/), or two points for fewer phonemes segmented (e.g., /tr/ /ik/). In addition to the three benchmark probes created for fall, winter, and spring, 20 additional alternate forms are available to monitor student progress. Recent data indicate that PSF in first grade is used to identify students who have not met the goal, and therefore the majority of technical adequacy information has been reported for PSF measures administered in kindergarten (DMG, 2008). Reliability for PSF ranges from .74 for single-probe administration to .90 for multi-probe administration, and alternate form reliability is .88. Additionally, predictive validity with the winter of first-grade NWF and spring of first-grade ORF is .62. Test-retest reliability data were also collected from a random sample of participating schools during each year of the Oregon Reading First project. Reliability data collected with random samples of 20 students in Grades K and 1 from eight randomly selected schools during the 2006-2007 school year produced reliability coefficients ranging from .60 to .75 for PSF.

DIBELS Nonsense Word Fluency (NWF)

DIBELS NWF (6th edition revised; Good & Kaminski, 2002) is a 1-minute, standardized, individually administered measure designed to assess students' acquisition of the alphabetic principle, or their knowledge of letter-sound correspondences and phonological recoding skills. During the 1-minute timing, students are presented with an 8.5" x 11" sheet of paper with consonant-vowel-consonant (CVC) and vowel-consonant (VC) words in which each letter represents its most common sound. During the creation of the measure, the difficulty of the words was determined by the word pattern (i.e., VC and CVC) and the relative difficulty of the consonants. In particular, the letters *q* and *x* are not used, since they typically represent more than one phoneme, the letters *h, w, y,* and *r* appear only in the word-initial position, and the letters *c* and *g* appear only in the word-final position. Additionally, real words and words that sounded like inappropriate words were excluded, but words that sounded like real words (e.g., sok) were not excluded (J. Wallin, personal communication, June 4, 2010). Once difficulty categories were identified, words were randomly ordered onto a page containing 10 rows of five nonsense words each.

The administration directions provide students with two options to approaching the reading task that account for varying skill levels; students are instructed to either provide each letter sound or to read the whole word. If the stimulus word is "sok," for example, students can respond either by saying /s/ /o/ /k/, or /sok/ and receive the same score of three correct letter sounds. Although the same score can be obtained via either of these word-reading approaches, because this is a fluency-based measure, students will

likely receive a higher score if they are phonologically recoding as opposed to producing the sounds in isolation.

Published alternate form reliability for this edition of first-grade NWF ranges from .83 to .96, and concurrent validity with ORF ranges from .69 to .78 and with the Woodcock Johnson Psycho Educational Battery-Revised ranges from .71 to .75. Test-retest reliability data were collected from a random sample of participating schools during each year of the Oregon Reading First project. Reliability data collected with random samples of 20 students in Grades K and 1 from four randomly selected schools during the 2008-2009 school year produced reliability coefficients ranging from .79 to .93 for NWF.

Modified Scoring Procedures for NWF

Literacy coaches working with school-building assessment teams were instructed to have their testers follow standard testing procedures by providing the standardized directions that accompany the measure, to explicitly mark how students approached each word-word using the "slashes and dashes" according to the *DIBELS Administration and Scoring Guide* (6th edition; Good & Kaminski, 2002), and to calculate two scores for the NWF measure: total Correct Letter Sounds (CLS) and Words Recoded Completely and Correctly (WRC). During quarterly DIBELS refresher trainings conducted prior to each benchmark data-collection period (i.e., fall, winter, and spring), literacy coaches were reminded of how critical it was that testers explicitly mark how the students approached the pseudo-words, as those markings would be used to categorize students' performance on each word into one of the four decoding strategies related to unitization. Using the

same unitization framework proposed by Harn et al. (2008), students' attempts on each

pseudo-word were placed into the decoding strategies outlined in Table 1.

TABLE 1. Word Reading Strategies Categorized
on DIBELS Nonsense Word Fluency

Strategy	Example
Sound-by-Sound	/s/ /o/ /k/
Word Recoded (sound-by-sound then recode)	/s/ /o/ /k/ /sok/
Partial Blending	/s/ /ok/
Whole Word Reading	/sok/

Trained research staff at the Oregon Reading First Center categorized student responses

based on the markings made by testers. To verify the consistency of categorization a

reliability check was conducted by having two staff members code 200 randomly

sampled probes from the 2008-2009 school year (50 probes each from Grades K and 1 in

winter and spring). Reliability coefficients ranging from .96 to .99 were obtained,

indicating strong agreement regarding staff understanding of how students decoded the

pseudo-words. A copy of the NWF probe designed to collect information about students'

decoding strategy use is provided in the Appendix.

In addition to this scoring modification, a third total score for NWF, the number

of words a student read as a whole unit correctly (WRWUC), was also calculated. To

receive credit on this scoring metric a student had to have attempted the word as a whole

unit without first attempting at the sound level or using a partial blend (both of which

would be considered a recode). Students could receive credit for one WRWUC if they

read the word as a unit and made an error but self-corrected that error within the 3-second

time limit set forth in the *DIBELS Administration and Scoring Guide* (6th edition; Good

91

& Kaminski, 2002). The rationale for calculating this score is based on research (Cummings et al., in press; Harn et al., 2008; Ritchey, 2008) that indicates that students who utilize a whole-word reading approach not only receive higher scores on NWF but, more important, read with significantly greater fluency on measures of ORF in both the winter and spring of first grade.

School-Level Measures

Indicators of socioeconomic status and students' initial academic achievement were included as school-level variables in these analyses. These particular variables were included as school context variables for three reasons: (a) Family socioeconomic characteristics and prior student achievement are two types of background variables that should be included in studies examining school effects (Teddlie et al., 2000); (b) they were similar to the criterion schools had to meet in order to be eligible in the Oregon Reading First initiative; and (c) research has indicated that even homogeneous schools with relatively similar contexts can account for significant proportions of observed variance in student achievement (Fien et al., 2009).

Socioeconomic Status

The first of these variables is the percentage of students eligible for free and reduced-price lunch (Percent FRL), a common index of socioeconomic status, as reported for each school by the National Center on Educational Statistics (NCES, 2010a). Although all schools were required to meet the poverty criterion of a minimum of 50% of the student population eligible for free and reduced-price lunches, examination of these

data reveal a not negligible amount of variability in the percentage of students eligible for free and reduced-price lunch (percentages ranged from 55.9% to 92.5%).

Academic Achievement

The second school-level variable of interest is the percentage of incoming kindergarteners categorized as being at risk for later reading difficulties (Percent K Risk). This variable was used throughout the Oregon Reading First project to represent the prior academic achievement of students and to determine the amount of technical assistance and on-site support schools might require to achieve the goal of all students reading proficiently in grade-level materials by the end of Grade 3. Additionally, research indicates that an increasingly common practice in school-effects research is to include an indicator of initial academic achievement to control for differences in prior attainment, thereby permitting a more accurate estimation of "value-added" by schools (Teddlie et al., 2000).

Procedures

During the 2008-2009 school year, schools collected first-grade fall benchmark data (i.e., PSF and NWF) from all students between August 25, 2008, and September 19, 2008. The first-grade DIBELS NWF spring benchmark probes were administered between April 27, 2009, and May 22, 2009. During the 2009-2010 school year, schools collected first-grade fall benchmark data (i.e., PSF and NWF) from all students between September 8, 2009, and September 25, 2009. The first-grade DIBELS NWF spring benchmark probes were administered between April 26, 2010 and May 21, 2010. School

assessment teams, trained by their literacy coach, collected the DIBELS data for each benchmark data-collection period. The literacy coach from each building was required to participate in a quarterly DIBELS refresher training prior to benchmark data collection to ensure that data were collected reliably and valid interpretations of student performance and expected to conduct a similar training with their assessment teams.

Analyses

The proposed research raises questions pertaining to student performance on NWF and characteristics of the words comprising the measure. Research Questions 1 and 2 specified models with performance on NWF as a function of (a) instructional support (i.e., school); and (b) student characteristics, respectively. Research Question 3 focused on the "difficulty" of words students are asked to read. Pseudo-words, treated as test items (coded 0,1), are hypothesized to be more or less difficult as a function of student-level characteristics and the school within which students were receiving instruction.

Nonsense words are apt to vary with respect to difficulty and may require different levels of skill to respond correctly. The proposed research tested hypotheses about the sources of variability in performance on NWF with regard to total Correct Letter Sounds (CLS) produced as well as Words Read as Whole Units Correctly (WRWUC). Although WRWUC is not an official score of the DIBELS NWF measure, research has indicated that the automaticity and fluency associated with being able to read words as whole units is related to performance on later measures of Oral Reading Fluency (Cummings et al., in press; Harn et al., 2008). Because Oregon Reading First

schools were charged with the responsibility of selecting reading curricula based on principles of scientifically based reading research that align with the five critical components of early reading identified by the National Reading Panel (NICHD, 2000) but were not required to adopt the same combinations of core, supplemental, and intervention programs, it was hypothesized that a significant amount of variance observed in students' CLS and WRWUC scores on NWF at the end of first grade would be explained by the schools within which they were enrolled.

Moreover, research has indicated that the following school context variables may also contribute significantly to differences observed in student performance: socioeconomic status and prior academic achievement (Teddlie et al., 2000). Additionally, because research has indicated that the level of knowledge and skill a student brings to the classroom can contribute significantly to later performance (Good et al., 1998; Hedges & Hedberg, 2007), it was hypothesized that students' initial level of skill will also account for a significant amount of variance observed in students' performance on NWF at the end of first grade. Also, as noted earlier, the rapidly increasing number of ELs in elementary schools provided a sound rationale for examining whether ELs obtained significantly different scores on DIBELS NWF in the spring of first grade compared to their English-speaking peers.

Multilevel Data Structure

For all research questions, the data had a multilevel structure. For Research Questions 1 and 2, students (Level 1) were nested in one of 14 elementary schools (Level

2). The data structure pertaining to Research Question 3 was more complex, as item responses (Level 1) were nested within students (Level 2) and students were nested in a school (Level 3).

Analyses of the multilevel data for Research Questions 1 and 2 required estimation of hierarchical linear models (HLM). The Hierarchical Linear Modeling (HLM) program, version 6.08 (Raudenbush, Bryk, Cheong, & Congdon, 2004) was used to conduct a logical sequence of model testing proceeding for more-to-less constrained parameter estimates and to test hypotheses associated with each research question.

Predictors of Performance on NWF (Research Questions 1 and 2)

Nonsense Word Fluency (NWF), total Correct Letter Sounds (CLS), and Words Read as Whole Units Correctly (WRWUC) were dependent variables with student characteristics and school context variables as predictors. Question 1 focused on the effect that schools with varying context (i.e., different student populations) may have had on students' performance on NWF. Research Question 2 examined the effects of student characteristics (i.e., students' initial level of performance on PSF and NWF in the fall of first grade and EL status) on students' performance on NWF. To examine the effects associated with the student- and school-level predictors, a two-level hierarchical linear regression model was estimated, with students at Level 1 and schools at Level 2.

All school-level predictors were grand-mean centered, adjusting the mean to represent the group mean value for a person with a (grand) average on each predictor; school-level predictors were grand-mean centered because 0 was not a meaningful value for either predictor, as all participating schools had a proportion of students eligible for

96

free and reduced lunch and some incoming kindergarten students identified as being at risk for later reading difficulties. For student-level predictors, students' scores on PSF and NWF were entered as raw scores because 0 was a meaningful value and student EL status (a dichotomous variable with 0 for non-English Learners and 1 for English Learners) was not centered, meaning that all coefficients in the models indicate the average predicted performance of non-ELs across participating schools.

The modeling process for both NWF outcomes of interest—CLS and WRWUC—began with an unconditional random intercepts model (Equations 1.1 and 1.2). Estimating this model provided information regarding the amount of between-school and within-school variation associated with student performance on NWF CLS and WRWUC effects, or the effect of school j on NWF for student i.

Level 1

$$NWF_{ij} = \beta_{0j} + r_{ij} \tag{1.1}$$

Level 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{1.2}$$

Because significant variation among students within school was observed, a second random-intercept model including student-level covariates was estimated. The student covariates included (a) students' initial performance on NWF in the fall of first grade, (b) students' initial performance on PSF in the fall of first grade, and (c) whether the student had been classified as an English language learner (ELL). Once a baseline model with all significant student-level predictors was established, school-level covariates—percentage of incoming kindergarten students at risk (Percent KRisk) and percentage of students eligible for free and reduced-price lunch (Percent FRL)—were

97

added to the model to determine if they accounted for any additional variance observed in CLS and WRWUC. The fully specified model for CLS and WRWUC (i.e., NWF) with all student- and school-level predictors is provided in equations 2.1 and 2.2 below.

Level 1

$$NWF_{ij} = \beta_{0j} + \beta_{1j}(\text{initial letter sounds}) + \beta_{2j}(\text{initial segmentation}) + \beta_{3j}(\text{ELL}) + r_{ij}$$

$$(2.1)$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(PercentKRisk) + \gamma_{02}(PercentFRL) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$

$$(2.2)$$

Item Characteristics (Research Question 3)

Research Question 3 pertained to item difficulty as a function of student-level characteristics. Multilevel regression models were estimated, testing how *item responses* (scored 0-incorrect, 1-correct) were influenced by the schools they attended during the 2008-2009 (Sample 1) and 2009-2010 (Sample 2) school years.

Hierarchical generalized linear models (HGLM) were used to determine if significant variability in item difficulties was evident in the first 20 pseudo-words on the spring of grade 1 benchmark probe. In theory, item difficulties should be non-significantly different because each VC and CVC pseudo-word is composed of letters making their most common letter sounds; no one combination of letter sounds should be significantly more difficult than others.

Though the Research Question focused specifically on item and student-level effects, it was necessary to structure the data with item responses nested within students nested within schools. The three-level data structure specified scored item responses (0,1) at Level 1, respondents at Level 2, and school at Level 3. Testing school effects on item difficulty required sequentially modeling item difficulties without school effects (Level 3) in the model, and subsequently with Level 3 in the model. It is important to note that all students who participated in this study responded to the same DIBELS NWF probe (spring of Grade 1; see Appendix). A student may not, however, have responded to all items because NWF measure is a 1-minute test of students' fluency and automaticity with letter-sound correspondences and decoding skills. To conduct the analyses without taking the fluency-based nature of the measure into consideration would be inappropriate because the difficulty of items that fewer students reached during the 1-minute timing could be overestimated. Therefore, to address this issue of fluency for Research Question 3, descriptive analyses were conducted to determine the number of students who read the first 20 pseudo-words on the spring NWF benchmark probe and the measure was truncated at that item. This particular item was selected not only because a majority of students reached that item (70%) but also because students only had to have responded correctly to these first 20 items (i.e., correctly produced the letter-sound correspondences in the first 20 VC and CVC words) to have met the benchmark goal for this measure, which is 50 CLS by the middle of first grade, to be considered on track for meeting later reading goals.

Testing the Effects of School on Item Difficulty

Examination of variability in item difficulty was tested with HGLM. Model testing proceeded from highly constrained to less constrained models. Initially, a two-level, intercept-only model (Model 1) was estimated (Equations 3.1, 3.2), where the dependent variable was the log-odds (logit) associated with the scored response (0,1) to each item 1 to (k-1) by person j, in school m. For this model, one item was not represented in the Level 1 model; it served as the intercept for the item coefficients.

Level 1

$$\log\left(\frac{p_{ijm}}{1-p_{ijm}}\right) = \beta_{0jm} + \beta_{1jm}X_{1jm} + \beta_{2jm}X_{2jm} + \cdots + \beta_{(k-1)jm}X_{(k-1)jm} \qquad (3.1)$$

Level 2

$$\begin{cases} \beta_{0jm} = \gamma_{00m} + u_{0jm} \\ \beta_{1jm} = \gamma_{10m} \\ \quad\vdots \\ \beta_{(k-1)jm} = \gamma_{(k-1)0m} \end{cases} \qquad (3.2)$$

Next, a 3-level model including school (Level 3) was estimated where the variance of log-odds (logit) associated with the scored response (0,1) to each item 1 to (k-1) by person j, estimated the variance explained by school m (Equations 3.1, 3.2, and 3.3).

100

Level 3

$$\begin{cases} \gamma_{00m} = \pi_{000} + r_{00m} \\ \gamma_{10m} = \pi_{100} \\ \quad\vdots \\ \gamma_{(k-1)0m} = \pi_{(k-1)00} \end{cases} \tag{3.3}$$

With addition of school (Equation 3.3), this model partitioned all item response

variability into four components: (a) variance among item responses within students,

(b) variance among students, (c) variance among schools, and (d) error. The item

difficulties were obtained as a function of the π coefficients. Specifically, the difficulty of

item i is $(-\pi_{i0} - \pi_{00})$.

CHAPTER IV

RESULTS

The effects of school- and student-level predictors on students' demonstration of the alphabetic principle, as well as the item difficulties for the first 20 pseudo-words on the spring of Grade 1 DIBELS NWF benchmark probe, are presented in this chapter. In the first section, descriptive statistics are provided for the Level 1 (student) and Level 2 (school) predictors. In the second and third sections, the results of the following models constructed to answer Research Questions 1 and 2 are reported for each outcome variable of interest, first for CLS and then for WRWUC: (a) unconditional Analysis of Variance (ANOVA) model, (b) single variable model with Level 1 predictors, (c) the baseline model with all significant Level 1 predictors, (d) single variable model with Level 2 predictors, and (e) the fully specified model with all significant Level 1 and Level 2 predictors. Although the number of school units is small ($N = 14$), full maximum likelihood (FML) estimation methods were found acceptable and used to allow for the calculation of hypothesis tests to examine the differences in model fit as a result of adding various fixed and random effects. In the fourth and final section, the results of the hierarchical generalized linear model (HGLM) and tests for differential item functioning among the first 20 items are presented.

Descriptive Statistics

Descriptive statistics for all student- and school-level predictors are reported in Table 2 below. Examination of the student-level predictors reveals that although the

102

mean scores for PSF and NWF in the fall of Grade 1 are greater than the benchmark goal

for that time of year (35 for PSF and 24 for NWF), a wide range of performance was also

evident on both of these measures, indicating that many students performed somewhat

below or above and significantly below or above the goal for benchmark performance.

Moreover, examination of the school-level descriptives reveals a relatively broad range

(49%) in the percentage of incoming kindergarten students identified as being at risk for

later reading difficulties, based on their performance on DIBELS measures in the fall of

kindergarten (i.e., Initial Sound Fluency and Letter Naming Fluency measures), and a

similarly broad range (36.6%) for the percentage of students eligible for free and

reduced-price lunch. Additionally, 26.6% of students in this sample were categorized as

English Learners (ELs) according to the following state guidelines: (a) Data obtained

from a home language survey indicated that English was not the primary language

spoken in the home, and (b) the student obtained a score at a level indicating he/she was

limited in his/her English language proficiency on a measure of English language

proficiency.

TABLE 2. Descriptive Statistics, Including Means, Standard Deviations, and Minimum
and Maximum Scores for School- and Student-Level Predictor Variables

Variable	N	M	SD	Minimum	Maximum
Level 2 Descriptive Statistics					
Percentage of incoming kindergarten students at risk	14	30.6	13.5	15.0	64.0
Percentage of students eligible for free and reduced-price lunch	14	71.2	12.0	55.9	92.5
Level 1 Descriptive Statistics					
PSF raw score (fall of Grade 1)	1,111	40.93	15.52	0	76
NWF raw score (fall of Grade 1)	1,111	30.94	22.05	0	147

Student-level descriptive statistics on the end-of-year outcomes for NWF in the

spring of first grade—CLS and WRWUC—are provided in Table 3. Examination of the

raw scores reveals that the average CLS score obtained by students in this study at the

end of first grade was 77.16, with a standard deviation of 39.41, while the average

WRWUC score was 18.93 words read as whole units correctly, with a standard deviation

of 16.89 during the 1-minute timing.

TABLE 3. Student-Level Descriptive Statistics, Including Means, Standard Deviations, and Minimum and Maximum Scores for DIBELS NWF Outcome Scores

Outcome	N	M	SD	Minimum	Maximum
Total Correct Letter Sounds	1,111	77.16	39.41	6	255
Words Read as Whole Units Correctly	1,111	18.93	16.89	0	84

Whereas student-level information is provided in Table 3, school-level descriptive

statistics for the NWF outcome sores are reported in Table 4; the scores reported here are

the grand mean across the schools for NWF CLS and WRWUC in the spring of Grade 1.

It is important to provide descriptive information about schools as well as students

because the analyses involved student-level information nested within schools. The mean

CLS score in the spring of Grade 1 for the 14 participating schools, for example, was

approximately 78 Correct Letter Sounds, with a standard deviation of 12.74, and the

mean WRWUC score was approximately 19 Words Read Correctly as Whole Units, with

a standard deviation of 5.24. Further examination of the minimum and maximum values

for each outcome variable of interest reveals a wide range in scores, 48.38 for CLS in the

spring of first grade and 20.71 for WRWUC, variability that may be due to inherent differences in student ability or potentially explained by specific student- and school-level factors.

TABLE 4. School-Level Descriptive Statistics, Including Means, Standard Deviations, and Minimum and Maximum Scores for DIBELS NWF Outcome Scores

Outcome	N	M	SD	Minimum	Maximum
Total Correct Letter Sounds	14	78.46	12.74	56.21	104.59
Words Read as Whole Units Correctly	14	19.18	5.24	6.00	26.71

Correlation of Student-Level Variables

Zero-order correlations reveal that students' fall of Grade 1 PSF score was minimally correlated with students' later performance on NWF in the spring of Grade 1, $r = .29$ with CLS and $r = .27$ with WRWUC. Not surprisingly, the correlations between CLS in the fall of Grade 1 with CLS and WRWUC in the spring were larger in magnitude, with correlations of $r = .63$ and $r = .57$, respectively. Additionally, students' fall of Grade 1 PSF and NWF scores were moderately correlated, $r = .42$ ($p < 0.01$). Although statistically significant, the correlations are not large and the substantive interest in these variables provides a sufficient rationale for including both predictors in the model.

Correlation of School-Level Variables

To ensure that the assumption of no multicollinearity between school-level predictors was supported, intercorrelations among the two school-level predictors of

interest—percentage of incoming kindergarten students at risk for later reading difficulties, percentage of students eligible for free and reduced-price lunch—were calculated. Results revealed a low, statistically insignificant correlation of $r = .38$, providing evidence of no multicollinearity.

Examining Observed Variance in CLS

To begin answering Research Questions 1 and 2 with respect to sound-level performance on NWF in the spring of Grade 1 (i.e., total CLS), an unconditional Analyses of Variance (ANOVA) model that included no student- or school-level predictors was employed to examine the magnitude of variation observed in Correct Letter Sounds (CLS) was attributed to between-school and within-school factors. More specifically, for the 14 schools that participated in the Oregon Reading First initiative and participating in this study, the unconditional ANOVA addressed the following research question: How much of the variance observed in CLS at the end of first grade exists between and within schools?

Unconditional ANOVA

The purpose of the unconditional ANOVA, or one-way ANOVA with random effects, is fourfold: (a) to provide an estimate of the average predicted CLS scores for the spring of Grade 1, (b) partition the total variance observed in those scores into variation that exists between and within schools, (c) provide a confidence interval for the mean CLS scores, and (d) test the hypothesis that significant variability exists between schools (Raudenbush & Bryk, 2002). Results of the fixed effects for unconditional ANOVA with

CLS as the outcome variable are reported in Table 5.

TABLE 5. Fixed and Random Effects From Unconditional Model
for Spring of Grade 1 NWF CLS Scores

Estimate of fixed effects				
Variable	β	*SE*	*t*	*df*
Predicted mean, γ_{00}	78.45	3.40	23.08*	13
Estimate of random effects				
Random effect	Variance	*SD*	χ^2	*df*
Mean μ_{0j}	140.79*	11.87	114.25*	13
Level 1 r_{ij}	1,424.53	37.74		

*$p < 0.001$.

These data indicate that the expected average CLS score for students in the spring of Grade 1 was approximately 78 correct letter sounds per minute, with the predicted range of scores being between 71.79 and 85.11 CLS. Additionally, calculations revealed that although the amount of variation that exists between schools appears small (9%), this variation is not only statistically significant and meaningful, but is consistent with findings from other, recent school effects research (Teddlie et al., 2000). Moreover, the results of this first model provide evidence that significant variation exists between schools for spring of Grade 1 CLS, and therefore it is appropriate to examine the contribution of student- and school-level characteristics in explaining this observed variation. In the subsequent sections, the single variable, baseline, and final models with significant student- and school-level predictors will be presented for CLS as the outcome of interest.

Single Predictor Models for CLS

As a preliminary stage of the model-building process, single predictor models

were built first, with only one student-level predictor entered at a time to determine the amount of variance contributed by each variable independently without controlling for any other variables. In particular, the percentage of variance accounted for was calculated by comparing the residual variance in the single variable model to the residual variance in the unconditional model. In each of these models, the three student-level predictors were entered as un-centered predictors because for fall of Grade 1 PSF and NWF, 0 is a meaningful score (that is, some students did have scores of 0 in the beginning of first grade on these two measures) and because entering EL status as an un-centered predictor permitted the examination of differences in performance between ELs and non-ELs on CLS. The results for each of the single variable models, with EL status, fall of Grade 1 PSF raw score, and fall of Grade 1 NWF raw score are reported in Table 6. Examination of these results reveals that, entered separately, student EL status, fall of Grade 1 PSF raw scores, and fall of Grade 1 NWF raw scores accounted for a statistically significant amount of the variance observed in the predicted average CLS score in the spring of Grade 1.

TABLE 6. Fixed and Random Effects for Single Predictor Models With CLS Means as Outcomes

	Estimate of fixed effects											
	EL status				Fall of Grade 1 PSF raw score				Fall of Grade 1 NWF raw score			
Variable	β	*SE*	*df*	*t*	β	*SE*	*df*	*t*	β	*SE*	*df*	*t*
Intercept	79.91	3.25	13	24.59**	48.36	4.32	13	11.18**	41.76	2.82	13	14.76**
Slope	-5.84	2.73	1,096	-2.14*	0.73	0.07	1096	10.10*	1.14	0.06	13	18.86**

	Estimate of random effects											
Random effect	Variance component	*SD*	*df*	χ^2	Variance component	*SD*	*df*	χ^2	Variance component	*SD*	*df*	χ^2
Mean μ_{0j}	120.72	10.99	13	108.83**	118.92	10.91	13	116.21**	74.09	8.61	13	47.09**
Mean μ_{1j}									0.03	0.16	13	26.08*
Level 1 r_{ij}	1419.69	37.67			1,304.58	36.11			829.65	28.80		

*$p < 0.05$.   **$p < 0.001$.

EL Status

The fixed coefficient for EL status was statistically significant, indicating that EL students, on average, could be expected to obtain a CLS score approximately 6 points lower than that of non-ELs in the spring of grade 1. In addition, adding EL status as a Level 1 predictor decreased the amount of variance observed in students' CLS scores. Calculation of the conditional intra-class correlation coefficient (ICC) for this model revealed that with the inclusion of EL status in the model the percentage of observed variance between schools (Level 2 units) decreased slightly to 7.8%, and calculation of the proportion reduction in variance at Level 2 indicated that 14.3% of the true between-school variance in CLS is accounted for by EL status.

Fall of Grade 1 PSF Raw Scores

Similarly, the inclusion of students' fall of Grade 1 PSF independently as a student-level predictor also resulted in a significant reduction in the amount of variance observed in CLS. Interestingly, although the fixed coefficient for the fall of Grade 1 PSF raw score was statistically significant, calculation of the conditional ICC for this model revealed that adding fall of Grade 1 PSF scores to the model resulted in no change to the percentage of observed variance between schools (9%). As noted earlier, however, the inclusion of this predictor resulted in a decrease in variance compared to the unconditional model with no predictors, from 140.79 to 118.92. Additionally, 15.2% of the true between-school variance in CLS was accounted for by fall of Grade 1 PSF scores.

Fall of Grade 1 NWF Raw Scores

Although student EL status and fall of Grade 1 PSF scores were entered as fixed predictors (i.e., the effects of EL status and fall of Grade 1 PSF scores were not allowed to vary randomly because to do so did not result in an improvement in model fit), allowing the effect of students' fall of Grade 1 NWF raw scores to vary randomly did result in a statistically significant improvement in model fit over its entry as a fixed predictor. Examination of the fixed coefficient from this model revealed that for students whose fall of Grade 1 CLS score was 0, every 1-point increase in students' fall of Grade 1 NWF score predicted a little over one correct letter sound increase in their spring of Grade 1 NWF score (β = 1.14).

Additionally, examination of the variance components compared to the unconditional model revealed that the variance associated with the model was almost reduced by half (140.79 - 74.09 = 66.7), and calculation of the conditional ICC revealed that by adding fall of Grade 1 NWF raw scores to the model 11.8% of the variance existed between schools. This was a slight increase compared to the unconditional model (9%) that was likely due to the shifting in variance between units, not poor model fit (as evidenced by the smaller values obtained for the variance components). Moreover, because the effect of fall of Grade 1 NWF raw scores was allowed to vary randomly for each student, it was possible to calculate the proportion of variance explained at Level 1; by adding fall of Grade 1 NWF raw scores as a predictor the amount of within-school variance observed in spring NWF CLS decreased by a substantial amount, 41.8%.

Although this was a substantial decrease in within-school variance, a significant amount of variance remained unaccounted for by this model. Furthermore, there is reason to believe that the relationship between critical early literacy skills is not always linear but may in fact be curvilinear. Recently conducted studies examining the relation between NWF and ORF have not only indicated that the relation may be curvilinear as opposed to linear (Fien et al., 2010) but also that growth on ORF in the primary grades may be curvilinear as opposed to linear where smaller gains are made in the upper grades (Chard et al., 2008; Fuchs, Fuchs, Hamlett, Walz, & German, 1993); a review of the literature revealed no studies examining the "shape" of growth for NWF. Although formal examination of growth was not possible in this study because only two time points were included, further examination of the correlation between fall and spring NWF CLS scores was nonetheless warranted. Therefore, to examine the possibility of this nonlinear relationship between fall of Grade 1 NWF scores and spring of Grade 1 CLS, a quadratic term was added to the model, producing the following Level 1 equation:

Spring of Grade 1 CLS = 34.92 + 1.55(NWF Raw) + (-0.004(NWF Squared)) + r_{ij}

This curvilinear relation between fall of Grade 1 observed NWF CLS scores and spring of Grade 1 predicted NWF CLS scores is presented in Figure 2. For illustrative purposes, the following values for the fall of Grade 1 CLS were selected: 0, *M*-1 *SD* (8.89), *M* (30.94), *M*+1 *SD* (52.99), and *M*+2 *SD* (75.04).
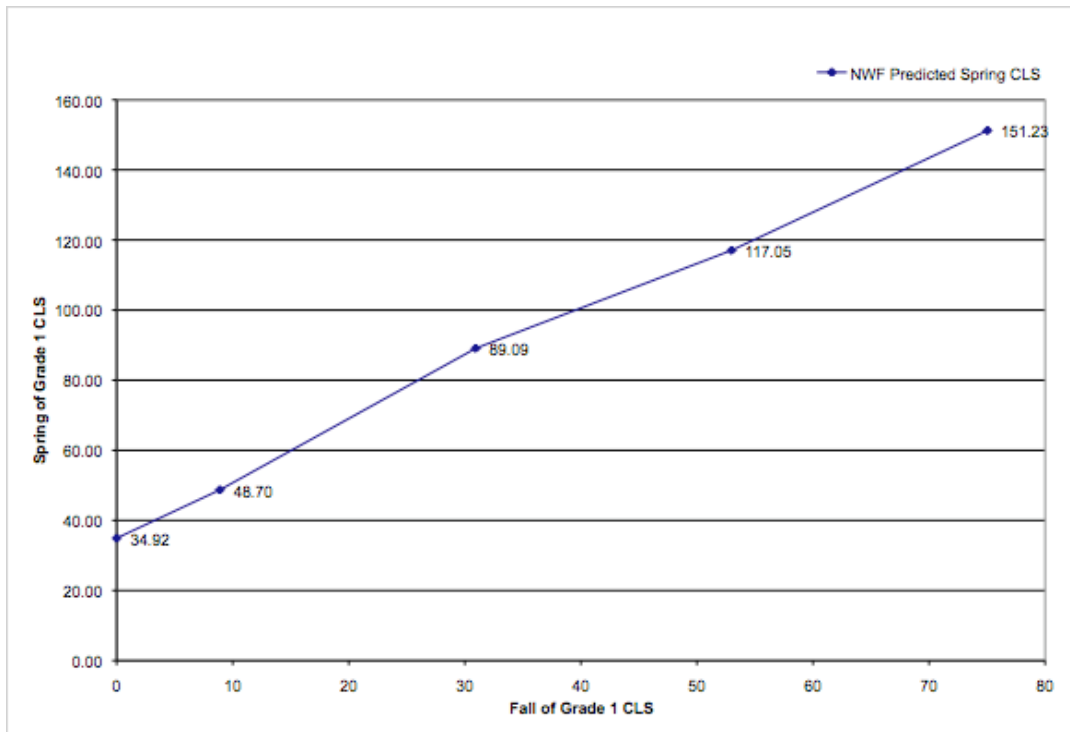
FIGURE 2. Curvilinear relationship between fall and spring of Grade 1 NWF
CLS scores.

Examination of this graph reveals, not surprisingly, a positive relation between

students' fall of Grade 1 CLS and spring of Grade 1 predicted CLS scores. Worth noting,

however, is the shape of this curvilinear relation. As students' fall of Grade 1 CLS scores

increase, the steepness of the slope begins to flatten slightly, indicating a "plateau effect"

that is also evidenced by the negative sign of the quadratic term. Although the curvilinear

relationship is slight (due to the small magnitude of the quadratic effect), it is possible

that this relation is observed as a result of a ceiling effect associated with measures that

has been documented for other early literacy measures (Harn et al., 2008; Paris, 2005). In

this case, a ceiling effect may be a byproduct of a constrained skill in which there are

only so many letter sound correspondences that can be accurately identified during one

minute, and/or the constrained measure used to measure students' knowledge of the

113

alphabetic principle that utilized only VC and CVC decodable words. It is possible, for example, that this "plateau effect" would not have been observed if a measure with more complex pseudo-words, such as multi-syllabic words, had also been included to provide a more comprehensive picture of students' alphabetic understanding and phonological recoding skills. Moreover, this relation may not be surprising in light of the potential diminishing returns in practical and instructional utility associated with higher levels of performance on NWF; previous studies have indicated, for example, that after a certain point, increased performance on NWF does not translate to higher scores on subsequent literacy measures such as ORF (Fien, et al., 2010).

Although calculation of the conditional ICC revealed a slight increase in the percentage of variation between schools (9.2% with the proposed model compared to 9% in the unconditional model), there were several indications that this model was the best fit for these data. Not only were the fixed coefficients for the NWF raw and squared terms statistically significant, but there was also a reduction in the amount of unexplained variance at Level 1 compared to previous models. Moreover, the results of a hypothesis test comparing the deviance statistics of this model with the previous model (with only the NWF raw score as a predictor) using full maximum likelihood (FML) to examine the change in model fit between the linear and nonlinear models was statistically significant, $\chi^2 (1) = 16.49$, $p < 0.001$. Lastly, calculations revealed that adding fall of Grade 1 NWF un-centered raw scores to the model and allowing the effect of those scores to vary randomly across students *and* including a quadratic term reduced the amount of within-school variation by 42.8%.

Before this could be designated as the baseline model for Level 1, it was

necessary to examine the amount of variance explained when all student-level predictors were included in the model to determine whether all predictors remained statistically significant. Including all student-level predictors diminished the effects of EL status and fall of Grade 1 PSF raw scores, which were no longer statistically significant once fall of Grade 1 NWF scores (raw and quadratic) were included in the model. Additionally, although the percentage of within-school variation was reduced to 42.8%, a comparison of the variance components between the two models revealed an increase in the variance components for the intercept and NWF slope. These results suggest that the previously described model with fall of Grade 1 NWF raw and quadratic scores is the best Level 1 model with the predictors available and therefore should be used as the baseline model for all subsequent analyses. The estimates obtained from the baseline model are reported in Table 7.

TABLE 7. Fixed and Random Effects for Baseline Model With CLS as Outcome

Estimate of fixed effects				
Variable	β	SE	df	t
Intercept γ_{00}	34.92	3.32	13	10.53**
NWF Raw Slope γ_{10}	1.56	0.11	13	13.83**
NWF Squared Slope γ_{20}	-0.004	0.001	1,082	-4.38**
Estimate of random effects				
Variable	Variance components	SD	df	χ^2
Mean μ_{0j}	82.47	9.08	13	52.78**
Mean NWF Slope μ_{1j}	0.02	0.15	13	25.69*
Level 1 r_{ij}	815.33	28.55		

*$p < 0.05$. **$p < 0.001$.

Establishing a Final Level 2 Model for CLS

Determining which multilevel model best fit these data required two steps: including the percentage of kindergarten students at risk (Percent K Risk) and percentage of students eligible for free and reduced lunch (Percent FRL) in the baseline model separately to test if either predictor was significant. If significant, then school-level predictors were included in the baseline model. The fixed coefficient for the Percent K Risk when added to the baseline model, for example, was statistically significant at the $p < .05$ level and reduced the proportion of variance at the school level by approximately 44.7%. When the Percent FRL was added as the only school-level predictor to the baseline model, that fixed coefficient was also statistically significant at the $p < .05$ level but produced an 8% increase in the variation associated with schools. Because each of the school-level predictors entered into the model individually were statistically significant, the next logical step was to build a model with both school-level predictors added to the intercept, resulting in a staggering reduction of 85.2% in the variance observed at Level 2.

A significant amount of variance in the intercept and NWF slope remained unexplained, however—$\chi^2 (11) = 21.64$, $p < .05$ for the intercept and $\chi^2 (13) = 25.90$, $p < .05$ for NWF slope—providing a reason to examine whether Percent K at risk and/or Percent FRL explained any of the variance observed in the slope. Percent K at risk was not a significant predictor of NWF slope, did not significantly reduce the amount of variance associated with the model, and did not produce a statistically significant improvement in model fit compared to the previous model with school-level predictors

on the intercept only, χ^2 (11), 3.02, $p > .05$. This was not the case, however, when the

Percent FRL was added to NWF slope; the results of this final model with all significant

student- and school-level predictors are reported in Table 8 and discussed in further detail

below.

TABLE 8. Fixed and Random Effects for Fully Specified 2-Level Model
With CLS as Outcome

Estimate of fixed effects				
Variable	β	*SE*	*df*	*t*
Intercept β_0				
Intercept γ_{00}	33.86	2.54	11	13.23***
Percent K At Risk γ_{01}	58.58	13.35	11	4.39**
Percent Eligible for FRL γ_{02}	-42.59	20.05	11	-2.12
NWF Raw Slope β_1				
Intercept γ_{10}	1.60	0.12	12	14.80***
Percent Eligible for FRL γ_{11}	-1.16	0.47	12	-2.45*
NWF Squared Slope β_2				
Intercept γ_{10}	-0.005	0.001	1,082	-4.76***

	Estimate of random effects			
Random effect	Residual Variance Component	*SD*	*df*	χ^2
Mean μ_{0j}	20.48	4.53	11	21.12*
Mean NWF Slope μ_{1j}	0.01	0.11	12	18.99
Level 1 r_{ij}	814.57	28.54		

*$p < .05$. **$p < .01$. ***$p < .001$.

In this model, the intercept value of 33.86 represented the predicted average

spring of Grade 1 CLS score for a student with a fall of Grade 1 NWF score of 0 enrolled

in a school with an average percentage of incoming kindergarten students at risk and an

average percentage of students eligible for free and reduced lunch. Dividing the value

obtained for the percentage of incoming kindergarten students at risk by 10 (58.55/10 =

5.85) indicates that for every 10% increase above in the percentage of incoming

kindergarten students at risk, students whose fall of Grade 1 score is 0 could be expected

to earn a spring of Grade 1 CLS score that is approximately 6 points *higher* than that of

their peers who also obtained a score of 0 CLS in the fall but attended schools with an

average proportion of incoming kindergarten students at risk (31.6%).

Dividing the value obtained for the percentage of students eligible for free and

reduced lunch by 10 (-42.59/10=-4.259) indicates that for every 10% increase above

average in the percentage of students eligible for free and reduced lunch, students whose

fall of Grade 1 score is 0 could be expected to produce between 4 and 5 letter sounds *less*

than their peers in schools with an average percentage of students eligible for free and

reduced lunch. This fixed effect, however, is no longer significant ($p = .057$) once the

proportion of students eligible for free and reduced lunch is added to NWF slope,

indicating that there is no significant correlation between the expected spring NWF CLS

score and Percent FRL for students whose fall of Grade 1 NWF CLS score was 0.

Despite this fact, this predictor was retained in the model because a hypothesis test

conducted comparing this model to a similar model where the percentage of students

eligible for free and reduced lunch was removed from the intercept revealed that the

deletion of this predictor did not result in a statistically significant improvement in model

fit, $\chi^2(1) = 3.21$, $p = 0.07$. On the contrary, a hypothesis test conducted comparing this

model to previous models with only one school-level predictor (Percent K Risk) indicated

a statistically significant improvement in model fit, $\chi^2(1) = 5.37$, $p < 0.05$. Furthermore,

this final model reduced the variance between schools in NWF slope to a non-significant

level, χ^2 (12) = 18.99, $p < 0.05$.

Significant Relation Between Percent FRL and CLS

The coefficient for the percentage of students eligible for free and reduced-price

lunch (-1.16) on NWF slope is significant, however ($p < 0.05$), indicating that for every

10% increase in the percentage of students eligible for free and reduced-price lunch,

students with a fall of Grade 1 CLS score of 0 can expect to produce between one and

two (1.48) fewer CLS in the spring of Grade 1. Although this difference appears trivial,

comparing the slopes for students with higher fall of Grade 1 CLS scores demonstrates

more clearly the magnitude of the predicted differences associated with students'

beginning of Grade 1 CLS scores. Illustrative data are reported in Table 9.

TABLE 9. Predicted NWF Slopes for Students in Schools With Different
Percentages of Students Eligible for Free and Reduced-Price Lunch

Fall of Grade 1 NWF CLS score	School with 20% less than average FRL (51.2%)	School with average FRL (71.2%)	School with 20% greater than average FRL (91.2%)
0	1.83	1.60	1.37
10	18.30	15.90	13.70
50	91.50	79.50	68.50

These data illustrate that although the differences in slope associated with the percentage

of students eligible for free and reduced-price lunch appear minimal in the results

obtained from the fully specified model, one has to remember that those values are the

predicted values for when students' fall of Grade 1 CLS score is 0. While there is only a

predicted difference in slope of 0.22 CLS for students' whose fall of Grade 1 NWF score

119

was 0 (1.83-1.37=0.22), comparison of the predicted slopes for students from schools with 20% less than average and 20% greater than average Percent FRL whose fall of Grade 1 NWF CLS scores were 50 CLS results in a predicted difference in slope of 23 CLS (91.50-68.50=23). Moreover, these data demonstrate that as Percent FRL for a school increases, NWF slope decreases, indicating that for students attending schools with above average Percent FRL, initial skill status (i.e., fall of Grade 1 NWF CLS scores) may be more important than for their peers in schools with a lower percentage of FRL.

Closer examination of the relation between students' fall NWF CLS scores and Percent FRL may, however, provide a viable explanation for these findings. This relation is depicted in the boxplots presented Figure 3 below, in which three categories of schools by Percent FRL are presented on the x-axis and students' fall NWF CLS scores are presented on the y-axis.
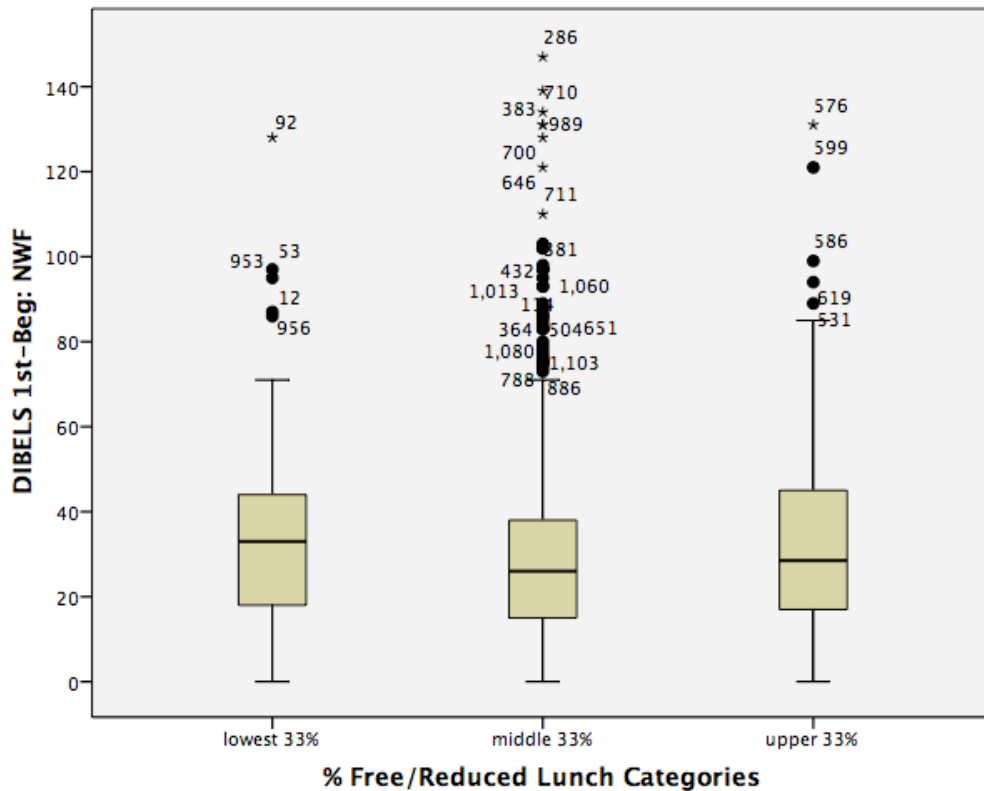
FIGURE 3. Relation between percent eligible for free and reduced-price lunch and fall NWF CLS scores.

In this figure, the 14 participating schools were grouped based on their Percent FRL into three groups: (a) one group of the three schools with the lowest Percent FRL (55.9% - 58.9%), (b) one group of the eight schools with a middle range of Percent FRL (59.8% - 81.2%), and (c) one group of the three schools with the highest Percent FRL (82.6% - 92.5%). Examination of these boxplots revealed, for example, that the means and standard deviations for fall CLS were fairly comparable (33.90 with a standard deviation of 20.16 for the lowest group, 29.99 with a standard deviation of 22.99 for the middle group, and 32.63 with a standard deviation of 22.52) and that the range of scores obtained for each Percent FRL was fairly large (0-147). Although the mean fall CLS

121

across Percent FRL contexts was comparable, it is worth noting that students attending schools with a lower Percent FRL had, on average, higher fall CLS scores and that the mean fall CLS score for students in these schools is almost 9 points above the benchmark goal of 25 CLS for the fall of grade 1. These high average fall CLS score may mean that students attending schools with a lower Percent FRL had less room to grow on NWF between fall and spring, thereby providing a potential explanation for the lower predicted spring CLS scores. Also, as noted earlier, previous research has indicated that there may be a point of diminishing returns on NWF where higher CLS scores only have so much practical value (Fien, et al., 2010) because the true purpose of reading is the ability to read connected text fluently and accurately for the purposes of comprehension.

Illustrating the Relation Between School-Level Predictors and CLS

These complex relations between the school-level and student-level predictors on the intercept and slopes associated with the model may best be understood, however, with some illustrative data. For this purpose, the fall of Grade 1 CLS values that will be used include 0, 8.89 (*M* − 1 *SD*), 30.94 (*M*), and 52.99 (*M* + 1 *SD*), and the predicted spring of Grade 1 CLS scores, intercepts, and slopes will be presented in two contexts for three example schools. In the first context, it is assumed that the average percentage of incoming kindergarten students at risk is held constant across all three schools, which vary in their percentage of students eligible for free and reduced-price lunch (20% less than the average, average, and 20% greater than average). In the second context, the manipulation of the school-level variables is reversed so that the average percentage of students eligible for free and reduced lunch is held constant while the percent of

incoming kindergarten students at risk for later reading difficulties is allowed to vary (20% less than average, average, and 20% greater than average). These illustrative data are presented in Tables 10 and 11.

Examining the predicted spring of Grade 1 NWF CLS scores, intercepts, and slopes for students with a range of scores in schools with an average percentage of incoming kindergarten students at risk and with three different percentages of students eligible for free and reduced lunch revealed three noteworthy trends in the data. First, students attending schools with fewer students eligible for free and reduced-price lunch had, on average, a higher predicted intercept in the spring of Grade 1 than their peers in schools with more students eligible for free and reduced lunch. Second, in each school presented, students with lower fall of Grade 1 NWF CLS scores had greater slopes than their peers in the same schools with higher predicted intercepts. Students with a fall CLS score of 0 attending schools with an average percentage of students eligible for free and reduced-price lunch, for example, had a predicted NWF slope of 1.60 CLS compared to their peers whose fall CLS score was one standard deviation above the mean (52.99) whose predicted NWF slope was only 1.09 CLS. Last, and perhaps most interesting, is the negative interaction observed between fall of Grade 1 CLS score and percentage of students eligible for free and reduced-price lunch. This negative interaction indicates that the correlation between the predicted spring of Grade 1 NWF CLS score and the Percent FRL becomes increasingly negative as fall of Grade 1 NWF CLS scores increase. That is to say, students with higher fall of Grade 1 NWF scores are expected to have *lower* spring of Grade 1 NWF CLS scores if they attend schools with higher than average percentages of students eligible for free and reduced-price lunch.

Examination of the data presented for schools where the percentage of students eligible for free and reduced-price lunch was held constant and the percentage of incoming kindergarten students identified as at risk for later reading disabilities was allowed to vary revealed some interesting trends as well. Similar to the previous example, students with the lowest fall of Grade 1 NWF CLS scores had the greatest slopes, regardless of school context. More interesting, however, is the prediction in the model that students attending schools with greater percentages of incoming at-risk kindergarteners will earn *higher* spring of Grade 1 NWF CLS scores than their peers with the same fall of Grade 1 NWF CLS scores in schools with a smaller percentage of incoming at-risk kindergarteners. First graders with an average fall of Grade 1 NWF score of 30.94 CLS attending schools with 20% greater than average incoming at-risk kindergarteners, for example, have a predicted spring of Grade 1 NWF CLS score of 90.51, a score 29.73 CLS higher compared to their peers in a school with 20% *less* than average incoming at-risk kindergarteners.

TABLE 10. Comparisons of Predicted Spring of Grade 1 CLS, Intercept, and Slope for Schools With Average, Less Than Average, and Greater Than Average Percentages of Students Eligible for Free and Reduced-Price Lunch

	20% less than average (51.2%)				Average percent FRL (71.2%)				20% greater than average (91.2%)			
	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*
Predicted CLS	42.38	58.29	94.50	126.06	33.86	47.71	78.80	105.24	25.34	37.12	70.28	84.41
Intercept	42.38	42.38	42.38	42.38	33.86	33.86	33.86	33.86	25.34	25.34	25.34	25.34
NWF Slope	1.83	1.75	1.54	1.33	1.60	1.51	1.30	1.09	1.37	1.28	1.07	0.86

Note. *M* = 30.94, *SD* = 22.05.

TABLE 11. Comparisons of Predicted Spring of Grade 1 CLS, Intercept, and Slope for Schools With Average, Less Than Average, and Greater Than Average Percentages of Incoming Kindergarten Students at Risk for Later Reading Difficulties

	20% less than average (10.6%)				Average percent K risk (30.6%)				20% greater than average (50.6%)			
	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*
Predicted CLS	22.14	35.99	67.08	93.52	33.86	47.71	78.80	105.24	45.58	59.42	90.51	116.95
Intercept	22.14	22.14	22.14	22.14	33.86	33.86	33.86	33.86	45.58	45.58	45.58	45.58
NWF Slope	1.60	1.51	1.30	1.09	1.60	1.51	1.30	1.09	1.60	1.51	1.30	1.09

Note. *M* = 30.94, *SD* = 22.05.

The results of this model indicate that the relation between Percent K Risk and Percent FRL and students' performance on DIBELS NWF is complex, but one that may be mediated by instruction. This mediation is evident in the predicted higher spring of Grade 1 NWF CLS scores for students attending schools with higher Percent K Risk, where the expectation might be that students attending schools with higher-risk populations of students would struggle more to acquire proficiency with early literacy skills. Additionally, the potential benefits associated with instruction are also supported by the fact that NWF slopes are larger for students at the greatest risk for later reading difficulties—those with a fall of Grade 1 NWF CLS scores of 0—for it is improbable that the improvement predicted in the model would happen independent of instruction.

Examining Observed Variance in WRWUC

To begin answering Research Questions 1 and 2 with respect to word-level performance on NWF in the spring of Grade 1 (i.e., WRWUC), an unconditional Analyses of Variance (ANOVA) model that included no student- or school-level predictors was employed to examine the magnitude of variation observed in Words Read as Whole Units Correctly (WRWUC) attributable to between-school and within-school factors. More specifically, for the 14 schools that participated in the Oregon Reading First initiative and participating in this study, the unconditional ANOVA addressed the following research question: How much of the variance observed in WRWUC at the end of first grade exists between and within schools?

The purpose of the unconditional ANOVA, or one-way ANOVA with random effects, is fourfold: (a) to provide an estimate of the average predicted WRWUC scores for the spring of Grade 1, (b) partition the total variance observed in those scores into variation that exists between and within schools, (c) provide a confidence interval for the mean WRWUC scores, and (d) test the hypothesis that significant variability exists between schools (Raudenbush & Bryk, 2002). Results of the fixed effects for unconditional ANOVA with CLS as the outcome variable are reported in Table 12.

TABLE 12. Fixed and Random Effects From Unconditional Model
for Spring of Grade 1 NWF WRWUC Scores

Estimate of fixed effects				
Variable	β	SE	t	df
Predicted mean, γ_{00}	19.19	1.34	14.31*	13
Estimate of random effects				
Random effect	Variance component	SD	χ^2	df
Mean μ_{0j}	21.31	4.62	96.70*	13
Level 1 effect, r_{ij}	265.36	16.29		

*$p < 0.001$.

These data indicated that the expected average CLS score for students in the spring of Grade 1 was approximately 19 whole words read per minute, with predicted scores ranging between 17 and 22 WRWUC. Additionally, calculations revealed that although the amount of variation that exists between schools appeared small (7.4 %), this variation was not only statistically significant and meaningful, but is consistent with findings from other, recent school effects research (Teddlie et al., 2000). Moreover, the

results of this first model provided evidence that significant variation exists between schools for spring of Grade 1 WRWUC, and therefore it is appropriate to examine the contribution of student- and school-level characteristics in explaining this observed variation. In the subsequent sections, the single variable, baseline, and final models with significant student- and school-level predictors will be presented for WRWUC as the outcome of interest.

<div align="center">Single Predictor Models for WRWUC</div>

Similar to the process for examining the amount of variance accounted for by the student- and school-level predictors of interest in performance on CLS, so too did the process for examining the contributions to WRWUC begin by entering one student-level predictor at a time to determine the amount of variance contributed by each predictor independently without controlling for any other predictors. The results of these analyses are reported in Table 13 and discussed in further detail below.

EL Status

When student EL status was entered as an independent predictor, a statistically significant fixed coefficient of -2.43 was obtained, indicating that EL students could be expected to read correctly, on average, two to three whole words *less* than their non-EL peers. Not only did the inclusion of this variable in the model result in a slight decrease of variance observed between schools (6.8% compared to 7.4% with the unconditional model), but the variance components around the intercept and Level 1 also decreased.

Furthermore, approximately 8.9% of the true between-school variance in the number of words students read correctly as whole units was accounted for by EL status.

Fall of Grade 1 PSF Raw Scores

Similarly, including students' fall of Grade 1 PSF raw scores as an individual predictor also produced a statistically significant fixed coefficient and resulted in a decrease of the variance components associated with the model. Although the variance associated with the intercept (spring of Grade 1 WRWUC scores) did not decrease much compared to the unconditional model (from 21.31 to 19.72), the decrease in variance associated with the residual was larger (from 265.36 to 246.73), indicating that the inclusion of this predictor in the model did reduce the amount of observed Level 1 variance. Furthermore, calculation of a conditional ICC to examine the amount of variance existing between schools indicated no significant improvement in model fit, as 7.4% of the variance is still associated with Level 2 units, but this is likely due to shifting of the variance between Level 1 and Level 2 units, not due to failure of fall of Grade 1 PSF raw scores to contribute significantly to the model. Lastly, additional calculations revealed that approximately 7.0% of the true between-school variance in the number of words students read correctly as whole units was accounted for by their fall of Grade 1 PSF raw scores.

TABLE 13. Fixed and Random Effects for Single Predictor Models With WRWUC Means as Outcomes

Estimate of fixed effects

Fixed effect	EL status				Fall of Grade 1 PSF raw score				Fall of Grade 1 NWF raw score			
	β	SE	df	t	β	SE	df	t	β	SE	df	t
Intercept	19.81**	1.37	13	14.44	7.33**	1.83	13	4.01	5.15**	1.03	13	5.01**
Slope	-2.43*	1.18	1,096	-2.05	0.29**	0.03	1096	9.16	0.44**	0.03	13	13.09**

Estimate of random effects

Random effect	Variance component	SD	df	χ^2	Variance component	SD	df	χ^2	Variance component	SD	df	χ^2
Mean μ_{0j}	19.41	4.41	13	89.90**	19.71	4.44	13	96.87**	7.05	2.65	13	31.38*
Mean μ_{1j}									0.01	0.11	13	41.22**
Level 1 effect, r_{ij}	264.61	16.27			246.72	15.71			172.80	13.15		

*$p < 0.05$.   **$p < 0.001$.

Fall of Grade 1 NWF Raw Scores

Lastly, a single variable model with students' fall of Grade 1 NWF raw score (i.e., CLS) as an un-centered, random predictor was built. Results of the model not only indicated that the coefficients for the intercept and NWF predictor were statistically significant but also that the inclusion of this predictor model resulted in a decrease of the variance associated with the model. The final estimation of the fixed effects for NWF revealed that for students' whose fall of Grade 1 NWF score was 0, every 1-point increase in fall of Grade 1 NWF CLS score could result in reading 0.44 more words as whole words correctly during the 1-minute timing. Although this finding is not readily interpretable, one could say that for every three additional CLS earned in the fall of Grade 1 students could, on average, be expected to read one more word correctly as a whole unit by the end of first grade.

The decision to allow the effect of fall of Grade 1 NWF raw scores to vary randomly between students was made based on the results of a hypothesis test comparing the deviance statistics between models where fall of Grade 1 NWF raw score was added as a fixed and random effect. Results of the hypothesis test indicated that allowing the effect of fall of Grade 1 NWF raw scores to vary randomly produced a statistically significant improvement in model fit, $\chi^2(2) = 18.57$, $p < .001$. Additionally, calculation of the conditional ICC for this model revealed that when fall of Grade 1 NWF raw scores were included the model, approximately 9.7% of the variance observed in WRWUC remained between schools, and by adding fall of Grade 1 NWF raw scores and allowing the effect of that raw score to vary randomly between students, the within-school observed variance was reduced by approximately 33.8%. Of all of the student-level

131

predictors, in other words, it appears that including fall of Grade 1 NWF raw scores in the model resulted in the greatest reduction of within-school variance. This finding may not be surprising in light of the fact that correct letter sound correspondence knowledge is required to read a pseudo-word correctly as a whole word.

<div align="center">Baseline Model for WRWUC</div>

Although the inclusion of fall of Grade 1 NWF raw scores resulted in a noticeable decrease in the variance associated with the model and a sizeable reduction in the amount of within-school variance, there was still a large portion of unexplained variance associated with the model. Based on the findings from the baseline model for CLS in which a curvilinear model best fit the data, a quadratic term (i.e., fall of Grade 1 NWF raw scores squared) was added to the previous model with only the NWF raw scores as a predictor, producing the following equation:

Spring of Grade 1 WRWUC = 2.73 + 0.61(NWF Raw) + (-0.002(NWF Squared)) + r_{ij}

This curvilinear relation between fall of Grade 1 observed NWF CLS scores and spring of Grade 1 predicted NWF WRWUC scores is presented in Figure 4. For illustrative purposes, the following values for the fall of Grade 1 CLS were selected: 0, *M*-1 *SD* (8.89), *M* (30.94), *M*+1 *SD* (52.99), and *M*+2 *SD* (75.04).
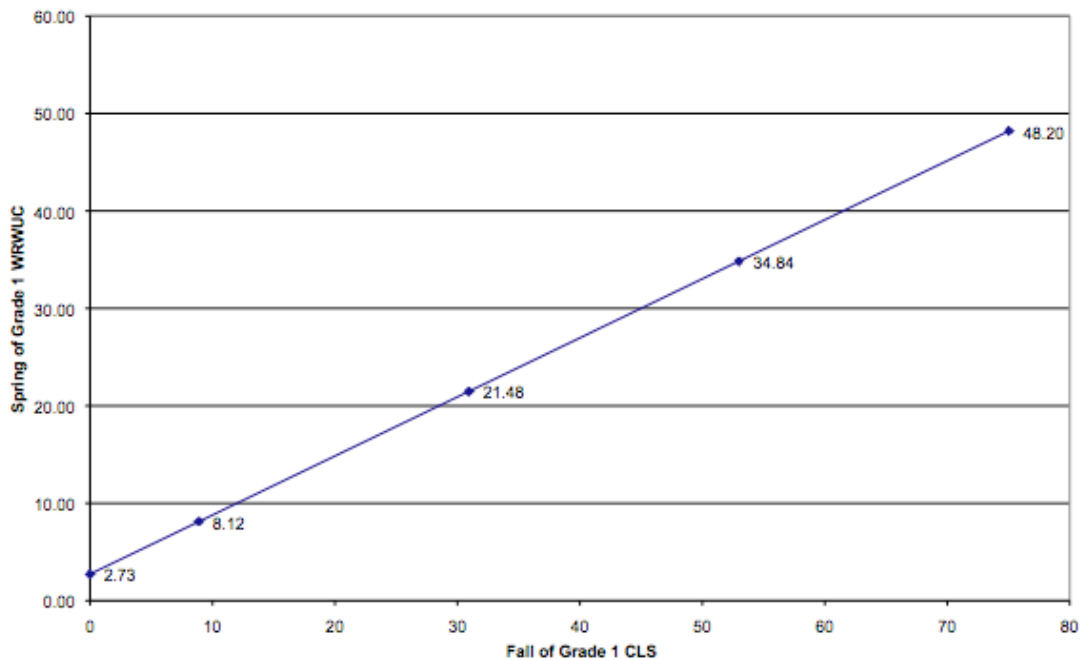
FIGURE 4. Curvilinear relation between fall of Grade 1 CLS and spring of Grade 1 WRWUC scores.

Results indicated that a curvilinear model was a statistically significant improvement over a linear model, as evidenced by statistically significant fixed effects for the linear and quadratic fall of Grade 1 NWF scores and a decrease in the variance components associated with the model. To determine whether the inclusion of the quadratic term resulted in a statistically significant improvement in model fit, a hypothesis test using FML to compare the deviance parameters of the linear and curvilinear models was conducted and a $\chi^2(1) = 16.49$, $p < .001$ was obtained, indicating a statistically significant improvement in model fit. Lastly, to evaluate overall improvements in model fit, a conditional ICC to determine the proportion of variance explained at the student level was calculated. The ICC indicated that with the inclusion of

133

the quadratic term to the previous model where the effect of fall of Grade 1 NWF un-centered raw scores was allowed to vary randomly between students, 4.2% of the variance remained between schools and that the within-school variance associated with WRWUC has been reduced by 35.8%.

Before deciding upon this as a baseline model for Level 1, it was necessary to examine the amount of variance explained when all student-level predictors were included in the model to determine which remained statistically significant. Including all student-level predictors diminished the effects of EL status and fall of Grade 1 PSF raw scores, which were no longer statistically significant once fall of Grade 1 NWF scores (raw and quadratic) were included in the model. Interestingly, even with these non-significant predictors, values obtained for the conditional ICC (4.2%) and within school variance accounted for by the model (35.8%) remained the same compared to the previous model with only significant NWF predictors. Comparison of the variance components associated with the intercept revealed a negligible decrease compared to the previous model (0.00617), which may be a result of small variability in WRWUC scores in the spring of Grade 1. Regardless, the previous model with only significant NWF predictors (raw and quadratic) was adopted as the baseline model because the other student-level predictors contributed minimally to the model once they were no longer significant. The results for the baseline model with only significant NWF predictors are reported in Table 14 below.

134

TABLE 14. Fixed and Random Effects for Baseline Model With WRWUC as Outcome

Estimate of fixed effects				
Variable	β	SE	df	t
Intercept γ_{00}	2.27	1.25	13	1.81
NWF Raw Slope γ_{10}	0.61	0.05	13	11.32**
NWF Squared Slope γ_{20}	-0.002	0.0004	1,082	-4.08**

Estimate of random effects				
Variable	Variance components	SD	df	χ^2
Mean μ_{0j}	7.47	2.73	13	33.48**
Mean NWF Slope μ_{1j}	0.10	0.009	13	39.69*
Level 1 r_{ij}	13.05	170.26		

*$p < 0.05$.   **$p < 0.001$.

Establishing a Final Level 2 Model for WRWUC

With a baseline model including student-level predictors established, the next step in the model-building process was to determine which, if any, of the school-level predictors (i.e., percentage of incoming kindergarten students at risk for later reading difficulties and percentage of students eligible for free and reduced lunch) accounted for significant amounts of the variance observed in the number of words students read correctly as whole units. This process entailed adding each school-level predictor to the baseline model independently to determine if the predictor was statistically significant and accounted for any of the observed variance in WRWUC, then adding all significant predictors to the intercept, and finally examining the effects of adding the significant school-level predictors to the slope of NWF. The results of the single Level 2 predictor models will be presented first, followed by the final Level 2 model with all significant predictors.

The first single Level 2 variable model built added the percentage of incoming kindergarten students at risk (grand-mean centered) to the previously established baseline model. Unfortunately, the inclusion of this predictor resulted in non-significant fixed effects for both the intercept (2.06, $p > 0.05$) and percentage of incoming kindergarten students at risk (13.74, $p > 0.05$), but the fixed effects for the linear and quadratic terms for fall of Grade 1 NWF raw scores remained statistically significant predictors. It is worth noting, however, that the inclusion of this school-level predictor in the model did result in a decrease of the variance components associated with the intercept and NWF slope and furthermore that the proportion of variance observed at the school level decreased by 40.1%.

The second single Level 2 variable model built included the proportion of students eligible for free and reduced lunch (grand-mean centered) to the baseline model. In this model, the fixed effects for the school- and student-level predictors were statistically significant, while the fixed effect for the intercept was not. Specifically, while the predicted average score of 2.25 WRWUC in the spring of Grade 1 was not significant, the expected 2.5 WRWUC decrease associated with every 10% increase in Percent FRL above the grand mean was significant. Furthermore, the inclusion of this predictor resulted in an increase in the variance associated with the slope and intercept and an increase (30.9%) of the variance associated with schools.

Based on the results of the final school-level model for CLS that included the percentage of incoming kindergarten students at risk and percentage of students eligible for free and reduced lunch on the intercept, it seemed logical to examine the same model with WRWUC as the outcome of interest, even though the proportion of incoming at-risk

kindergarten students was not, by itself, a statistically significant predictor. Examination of this model with all school-level predictors revealed not only that the fixed effects of *both* school-level *and* student-level predictors were statistically significant at the $p < 0.01$ level, but also that the inclusion of both predictors in the model greatly reduced the amount of variance associated with the intercept (7.47 for the baseline model, 1.35 for the new model with school-level predictors). Additionally, a hypothesis test conducted comparing this model to the previous single Level 2 variable models indicated that this model with both Level 2 predictors produced a statistically significant improvement in model fit, χ^2 (2) = 9.45, p = .009, and calculations revealed a staggering 81.7% reduction in the proportion of variance in WRWUC observed between schools.

A significant amount of variance in the intercept remained unexplained, however, χ^2 (11) = 21.93, p < .05, providing a reason to examine whether the percentage of incoming kindergarten students at risk and/or the percentage of students eligible for free and reduced-price lunch explained remaining variance observed in the slope. These next steps revealed that including the percentage of incoming kindergarten students at risk was not a significant predictor of NWF slope, did not significantly reduce the amount of variance associated with the model, and did not produce a statistically significant improvement in model fit compared to the previous model with school-level predictors on the intercept only, χ^2 (1), 3.15, p = .07. This was not the case, however, when the percentage of students eligible for free and reduced-price lunch was added to NWF slope; the results of this final Level 2 model (this is the final model because no school-level predictors remain and cannot be allowed to vary randomly) are reported in Table 15 and discussed in further detail below.

TABLE 15. Fixed and Random Effects for Fully Specified 2-Level Model
With WRWUC as Outcome

	Estimate of fixed effects			
Fixed Effect	β	*SE*	*df*	*t*
Intercept β_0				
Intercept γ_{00}	1.83	1.09	11	1.68
Percent K At Risk γ_{01}	16.46	5.93	11	2.78*
Percent Eligible for FRL γ_{02}	-11.92	8.36	11	-1.43
NWF Raw Slope β_1				
Intercept γ_{10}	0.64	0.05	12	13.14**
Percent Eligible for FRL γ_{11}	-0.77	0.20	12	-3.88**
NWF Squared Slope β_2				
Intercept γ_{10}	-0.002	0.0004	1,082	-4.60***

	Estimate of random effects			
Random effect	Residual Variance Component	*SD*	*df*	χ^2
Mean μ_{0j}	2.36	1.53	11	20.37*
Mean NWF Slope μ_{1j}	0.001	0.04	12	19.13
Level 1 r_{ij}	170.33	13.05		

*$p < .05$.   **$p < .01$.   ***$p < .001$.

In this model, the intercept value of 1.82 represented the predicted average spring of Grade 1 WRWUC score for a student with a fall of Grade 1 NWF score of 0 enrolled in a school with an average percentage of incoming kindergarten students at risk and an average percentage of students eligible for free and reduced lunch. Dividing the value obtained for the percentage of incoming kindergarten students at risk by 10 (16.46/10 = 1.65) indicates that for every 10% increase in the percent of incoming kindergarten students at risk, a student whose fall of Grade 1 score is 0 could be expected to earn a spring of Grade 1 WRWUC score between 1 and 2 points *higher* than that of their peers

who also obtained a score of 0 CLS in the fall but attended schools with an average percentage of incoming kindergarten students at risk (31.6%). Dividing the value obtained for the percentage of students eligible for free and reduced lunch by 10 (-11.92/10 = -1.92) indicates that for every 10% increase above average in the percentage of students eligible for free and reduced lunch, students whose fall of Grade 1 score is 0 could be expected to produce approximately two words read correctly as whole units less than their peers in schools with an average percentage of students eligible for free and reduced lunch. This fixed effect, however, is no longer significant ($p > .05$) once the percentage of students eligible for free and reduced lunch is added to NWF slope. Despite this fact, this predictor was retained in the model because a hypothesis test conducted comparing this model to a similar model where the percentage of students eligible for free and reduced lunch was removed from the intercept revealed that the deletion of this predictor did not result in a statistically significant improvement in model fit, χ^2 (1) = 3.21, $p < .05$. The coefficient for the percentage of students eligible for free and reduced-price lunch (-0.77) on NWF slope is significant, however, $p < .05$, indicating that for every 10% increase above average in the percentage of students eligible for free and reduced-price lunch students with a fall of Grade 1 CLS score of 0 can expect to produce 0.77 fewer WRWUC in the spring of Grade 1.

Illustrating the Relation Between School-Level Predictors and WRWUC

These complex relations between the school-level and student-level predictors on the intercept and slopes associated with the model may best be understood, however, with some illustrative data. For this purpose, the following fall of Grade 1 CLS values that will

be used include 0, 8.89 ($M - 1$ SD), 30.94 (M), and 52.99 ($M + 1$ SD), and the predicted

spring of Grade 1 WRWUC scores, intercepts, and slopes will be presented in two

contexts for three example schools. In the first context, it is assumed that the average

percentage of incoming kindergarten students at risk is held constant across all three

schools, which vary in their percentage of students eligible for free and reduced-price

lunch (20% less than the average, average, and 20% greater than average). In the second

context, the manipulation of the school-level variables is reversed so that the average

percentage of students eligible for free and reduced lunch is held constant while the

percentage of incoming kindergarten students at risk for later reading difficulties is

allowed to vary (20% less than average, average, and 20% greater than average). These

data are presented in Tables 16 and 17.

TABLE 16. Comparisons of Predicted Spring of Grade 1 WRWUC, Intercept, and NWF Slope for Schools With Average, Less Than Average, and Greater Than Average Percentages of Students Eligible for Free and Reduced-Price Lunch

	20% less than average (51.2%)				Average percent FRL (71.2%)				20% greater than average (91.2%)			
	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*
Predicted WRWUC	4.21	11.07	26.65	40.19	1.83	7.32	19.50	29.64	-0.56	3.56	17.11	19.09
Intercept	4.21	4.21	4.21	4.21	1.83	1.83	1.83	1.83	-0.56	-0.56	-0.56	-0.56
NWF Slope	0.79	0.75	0.66	0.57	0.64	0.60	0.51	0.42	0.48	0.44	0.35	0.26

Note. *M* = 30.94, *SD* = 22.05.

TABLE 17. Comparisons of Predicted Spring of Grade 1 WRWUC, Intercept, and NWF Slope for Schools With Average, Less Than Average, and Greater Than Average Percentages of Incoming Kindergarten Students at Risk for Later Reading Difficulties

	20% less than average (10.6%)				Average percent K risk (30.6%)				20% greater than average (50.6%)			
	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*	0	*M*-1 *SD*	**M**	*M*+1 *SD*
Predicted WRWUC	-1.46	4.03	16.21	26.35	1.83	7.32	19.50	29.64	5.12	10.61	22.79	32.93
Intercept	-1.46	-1.46	-1.46	-1.46	1.83	1.83	1.83	1.83	5.12	5.12	5.12	5.12
NWF Slope	0.64	0.60	0.51	0.41	0.64	0.60	0.51	0.41	0.64	0.60	0.51	0.41

Note. *M* = 30.94, *SD* = 22.05.

Examining the predicted spring of Grade 1 NWF WRWUC scores, intercepts, and slopes for students with a range of scores in schools with an average percentage of incoming kindergarten students at risk and with three different percentages of students eligible for free and reduced lunch reveals three noteworthy trends in the data. First, across the three school contexts presented, students with lower fall of Grade 1 CLS scores had higher slopes than their peers who displayed greater proficiency with letter-sound correspondences. Second, students attending schools with smaller percentages of students eligible for free and reduced lunch had higher predicted intercepts than their similarly performing peers in schools with higher percentages of students eligible for free and reduced-price lunch. Although the model predicted, for example, that students attending a school with 20% above average FRL will have an intercept of -0.56, this score is not possible and should be interpreted as predicting that students in these schools are predicted to be less likely to produce any WRWUCs in the spring of Grade 1. Last, students attending schools with higher percentages of students eligible for free and reduced-price lunch are predicted to have a lower WRWUC score than their similarly performing peers in schools with lower percentages of FRL. Students, for example, who had a mean CLS score of 30.94 in the fall of Grade 1 attending a school with 51.2% FRL are predicted to read between 26-27 words as whole units in the spring of Grade 1 compared to their peers attending a school with 71.2% FRL, who are predicted to produce only 17 words as whole units, a 10-point difference that is instructionally meaningful and due solely to differences in schools' socioeconomic context.

Examination of the data presented for schools where the percentage of students eligible for free and reduced-price lunch was held constant and the percentage of

incoming kindergarten students identified as at risk for later reading disabilities was allowed to vary reveals some interesting trends as well. Similar to the previous example, students with the lowest fall of Grade 1 NWF CLS scores had the largest slopes, regardless of school context, which may be expected because students with lower scores have more room to improve than students who began the year with higher CLS scores and who were potentially already reading words as whole units. It is worth noting that the predicted slopes remain constant across the three school contexts because the percentage of kindergarten students at risk was not a significant predictor of NWF slope, so across contexts it is predicted that students with lower scores will have greater slopes. Second, and perhaps more interesting, is the finding that students attending schools with higher percentages of at-risk kindergarteners are predicted to have higher WRWUC scores in the spring of Grade 1 than their similarly performing peers in schools with lower percentages of at-risk kindergarteners. A student with an average fall NWF CLS score (30.94) attending a school with only 10.6% of at-risk kindergarteners is predicted to read only 16.2 words as whole units by the spring of first grade compared to a similarly performing student attending a school with 50.6% (or average) Percent K at risk, who is predicted to read 22.72 words as whole units, a difference that is instructionally meaningful and may have practical implications for teachers.

Examining Variability of Item Difficulty in NWF

The last research questions guiding this study were designed to examine whether there was significant variability in the item difficulties for the pseudo-words on the spring of Grade 1 NWF benchmark probe. Of particular interest were the following research

143

questions:

1. Do the difficulties of the first 20 VC and CVC nonsense words on the spring of Grade 1 NWF benchmark probe differ significantly?

2. Is there a predictive relationship between (a) students' EL status, (b) fall of Grade 1 PSF raw scores, and (c) fall of Grade 1 NWF CLS scores and their spring of Grade 1 CLS score?

To answer these research questions, a single-factor Confirmatory Factor Analysis, equivalent to a multilevel, one-parameter logistic Rasch model, was conducted to examine the variability of item difficulties and whether any of the observed variability could be accounted for by student-level characteristics while controlling for school effects. This approach was used because multilevel modeling allows for an appropriate consideration of the nested structure of the data (i.e., items within students within schools) and an estimation of the difficulty of each individual item (or pseudo-word) on the NWF measure. In particular, Rasch analyses are designed to provide estimates of item difficulty in terms of the odds that a student with given ability will respond correctly. In the logit scale, an item scale value of zero indicates that a student with average ability will have a 50% chance of responding to the item correctly (Embretson & Riese, 2000).

For these data, a student with average alphabetic understanding and phonological recoding skills would have a 50% chance of correctly identifying the VC or CVC words on the measure. Items with negative difficulties are easier, indicating that students have a greater likelihood of correctly identifying all of the letter sounds in the word, while items with positive difficulties are more difficult, indicating that students with an average understanding of letter-sound correspondences and phonological recoding abilities will

have a decreased probability of identifying all of the letter sounds in the word correctly.

A series of increasingly complex models were built to ensure that the final model included all significant predictors; the results of the unconditional model with no student-level predictors are presented first, followed by the results of models with significant student-level predictors.

Unconditional Model

As noted earlier, purposes of the unconditional model include (a) to determine whether sufficient, significant variance exists that warrants further investigation (i.e., can that variance be explained by any additional predictors?); and (b) to partition the variance observed into between-school and within-school factors. To this end, answering the first research question related to the possibility of variability in item difficulties required the construction of an unconditional, one-parameter logistic Rasch model with the first 20 pseudo-words of the spring of Grade 1 NWF benchmark probe. As noted in Chapter III, the first 20 words were modeled for three reasons: (a) to control for the first fluency-based nature of the model (i.e., because not all students read all of the pseudo-words on the probe during the 1-minute timing, the difficulties for those items not reached by many students are likely to be overestimated); (b) students only had to correctly identify the letter-sound correspondences in the first 20 words to achieve the benchmark goal of 50 CLS and be considered on-track for later reading outcomes; and (c) 70.5% of students in the sample read at least 20 pseudo-words during the 1-minute test administration. Using Mplus (Múthen & Múthen, 2010) allowed for a direct estimation of item difficulties with the most accurate estimation methods, which will be reported later in conjunction with

the item difficulties estimated from the conditional model.

Examination of the results from the unconditional model revealed that the Rasch-estimated difficulties for the pseudo-words ranged from -3.857 to -0.364, indicating that, at the end of first grade, students with average levels of alphabetic understanding and phonological recoding skills had at least a 50% chance of accurately identifying *all* of the letter sounds that appeared in each of the first 20 nonsense words on the measure. According to these results, the easiest pseudo-word for students in the sample was *ut*, which may be due to the fact that it is a common letter combination that appears as the rime in many decodable CVC words taught in the early grades (e.g., h-ut, g-ut, n-ut, r-ut, etc.). In contrast, the most difficult word was *rec*, perhaps because most words in English that end with the letter *c* are preceded by *i,* not *e* (e.g., acidic, acerbic, agnostic, etc.).

Three additional findings associated with the model warrant further discussion as well. First, examination of the *p*-values associated with each item difficulty indicates that the individual items do vary substantially in their difficulty for students. Second, item difficulty did not increase uniformly throughout the measure, so items in the beginning of the measure are not necessarily easier for students with average alphabetic understanding than those at the end of the measure. Third, calculation of the ICC associated with the model revealed that approximately 32.3% of the observed variance in student performance on individual items was attributable to variation between schools.

Predictive Ability of EL Status and Prior PSF and NWF Scores

To examine the second research question related to variability in item difficulties,

146

a series of two-level measurement Rasch models with several combinations of student-level predictors were conducted. More specifically, single predictor models with each student-level predictor entered individually were conducted, and because all three were independent statistically significant predictors of student performance, the decision was made to create a fully specified model with all student-level predictors. In this model two of the three student-level predictors—students' EL status and fall of Grade 1 NWF scores—were significant predictors of how students performed on the first 20 pseudo-words in the spring of Grade 1. These results indicate that, on average, EL students were estimated to have a lower score by -0.327 logits or, alternatively, that native English speakers were 1.39 times more likely to provide a correct response than EL students. Additionally, results indicated that for every correct letter sound read in the fall students increase their chances of getting an item (i.e., pseudo-word) correct in the spring by 0.021 on the logit scale compared to students with a fall CLS score of 0. The results of this model are reported in Table 18 and discussed in further detail below.

TABLE 18. Results of 2-Level HGLM With All Significant
Student-Level Predictors

	Estimate of fixed effects		
Fixed effect	β	*SE*	*p*
Intercept			
EL status	-0.327	0.132	0.013
Fall of grade 1 NWF	0.021	0.002	< 0.001
	Estimate of random effects		
Random effect	Variance	*SD*	*p*
Mean μ_{0j}	0.225	0.128	0.08
Level 1 r_{ij}	1.273	0.08	< 0.001

The coefficient of -0.327 associated with EL status indicates that, on average, EL students were estimated to have a lower score by 0.327 logits, or that native English speakers were 1.39 times more likely to provide a correct response than EL students. Additionally, the coefficient of 0.021 associated with fall of Grade 1 NWF indicates that for each correct letter sound read in the fall, a student increases his/her chances of getting an item correct in the spring by a value of 0.021 on the logit scale, compared to a student who read no correct letter sounds. In other words, a student who read the average number of letter sounds correctly in the spring (77.16) was approximately 2.2 times more likely to get each item correct compared to a student who was one standard deviation below the mean (37.75), and 5.0 times as likely to do so as a student who read no correct letter sounds.

The item difficulties obtained from the unconditional and conditional models are provided in Table 19 for comparative purposes. Cursory examination of these difficulties indicates that the estimated difficulties of far fewer items in the conditional model are statistically significant compared to the unconditional model

TABLE 19. Rasch-Estimated Difficulty of 20 VC and CVC Nonsense Words
From the Unconditional and Conditional Models

Item Number	Pseudo-Word	Unconditional Model			Conditional Model		
		Item Difficulty	SE	p	Item Difficulty	SE	p
1	pov	-0.998	0.22	< 0.001	-0.096	0.21	0.649
2	riz	-2.407	0.23	< 0.001	-1.443	0.26	< 0.001
3	hal	-1.404	0.23	< 0.001	-0.546	0.27	0.004
4	jok	-0.811	0.17	< 0.001	0.079	0.22	0.719
5	bel	-1.517	0.17	< 0.001	-0.659	0.23	0.004
6	ib	-1.478	0.16	< 0.001	-0.602	0.20	0.003
7	tum	-2.77	0.21	< 0.001	-1.838	0.26	< 0.001
8	kaj	-0.951	0.16	< 0.001	-0.070	0.19	0.713
9	rec	-0.364	0.16	0.019	0.523	0.22	0.019
10	ut	-3.857	0.33	< 0.001	-2.692	0.39	< 0.001
11	nej	-1.142	0.20	< 0.001	-0.238	0.23	0.309
12	hos	-1.006	0.23	< 0.001	-0.095	0.22	0.665
13	um	-3.314	0.22	< 0.001	-2.397	0.21	< 0.001
14	waf	-0.505	0.15	0.001	0.388	0.20	0.054
15	joz	-0.799	0.22	< 0.001	0.118	0.24	0.627
16	kef	-1.201	0.19	< 0.001	-0.301	0.23	0.187
17	hod	-1.106	0.21	< 0.001	-0.180	0.25	0.463
18	dek	-1.317	0.20	< 0.001	-0.424	0.23	0.070
19	eb	-0.808	0.16	< 0.001	0.095	0.18	0.589
20	lad	-0.735	0.16	< 0.001	0.178	0.23	0.446

Note: The item difficulties for the unconditional and conditional models are not directly comparable because they were estimated with different structural models.

In addition to noting that the majority of item difficulties in the conditional model are no longer significantly different from 0 (12/20, or 60% of items), two additional points warrant further discussion. First, comparison of the item difficulties presented in

Table 19 shows that the difficulties for all 20 items were lower (i.e., easier) in the unconditional model than in the conditional model with student-level predictors, although many of these higher difficulties were not statistically significant. Furthermore, while all of the item difficulties in the unconditional model were less than 0, indicating a likelihood that all students, even those with lower alphabetic understanding skills, could respond correctly to each pseudo-word, this was not the case for the conditional model. Some of the item difficulties in the conditional model were greater than 0, indicating that these items were more challenging, even for students with greater alphabetic understanding and there was not a 50% chance that all students would respond to all items correctly. It is worth noting, however, that the numerical value for item difficulty associated with this analysis is relative and that interpretation of how one item differs from another (from easy to challenging) depends solely on the relative scale produced by this particular sample of students.

Second, and of more theoretical interest, is that the item difficulties obtained by both the unconditional and conditional models did not fully coincide with the claims made in the reading literature that words beginning with stop sounds (e.g., /p/, /b/, /d/, /k/, and /t/) are more difficult for students to read than words beginning with continuous sounds (e.g., /h/, /j/, /r/, /n/, and /w/; Chard & Osborn, 1999; Juel, 1991; Kame'enui & Simmons, 1990; Smith et al., 1998). Examination of the item difficulties obtained from both models revealed that some items beginning with a stop sound had lower item difficulties than several of the items that began with a continuous sound. The implications of this finding will be discussed in the next chapter.

Although these differences in item difficulties are interesting and provide

evidence of variability in item difficulties, the following steps were taken to further examine the contribution of the student-level predictors: (a) comparison of model fit statistics for the unconditional and conditional models, and (b) calculation of the ICC for the conditional model. The fit statistics from the unconditional and conditional models appear in Table 20.

TABLE 20. Comparison of Model Fit Statistics for Multilevel Rasch Models

Fit Statistic	Unconditional Model	Conditional Model
-2 log likelihood	-10090.798	-10041.311
AIC	20263.596	20132.621
BIC	20469.129	20178.540

Comparison of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) revealed a decrease in the information criterion from the unconditional to the conditional models, indicating that the conditional model with student-level predictors was better suited to the data. Additionally, comparisons of the -2 log likelihood values obtained from the model showed a statistically significant improvement in model fit, χ^2 (14) = 49.487, $p < 0.001$, and examination of the residual variances associated with each model also revealed an improvement in model fit as evidenced by a reduction in those variances (1.568 - 1.273 = 0.295). Calculation of an ICC for the conditional model also revealed a reduction in the variance associated with Level 2 units, 27.9% compared to 32.3% for the unconditional model. Lastly, examination of the variance components associated with the conditional model revealed that the inclusion of these student-level predictors in the model reduced the variance between schools to a non-significant level and therefore explained the differences observed in the item difficulties.

151

Summary

In summary, the purpose of this study was twofold: (a) to examine the school-level and student-level predictors of first-grade students' demonstration of the alphabetic principle via their accuracy with letter-sound correspondences and phonological recoding of VC and CVC nonsense words, and (b) to examine the structure of the measure itself to determine whether each VC and CVC word is of comparable difficulty for students with average alphabetic understanding. The same model-building process was followed to examine the variation in CLS and WRWUC. Initially, to answer the research questions focused on school- and student-level predictors, unconditional analyses of variance (ANOVAs) with no student- and school-level predictors were conducted to examine the magnitude of variation in Nonsense Word Fluency performance attributed to between- and within-school factors (Raudenbush & Bryk, 2002). Because the unconditional ANOVAs indicated that significant variation existed between schools, the following model-building process was used to further examine the existing variation: (a) single-variable models to examine the amount of variation accounted for by each student-level predictor independently, (b) a baseline model with all significant student-level predictors, (c) single-variable Level 2 models in which school-level predictors were added to the established baseline model to determine their significance, and (d) a fully specified model with all significant student- and school-level predictors to examine not only the within-school variation predicted by students' fall of Grade 1 NWF CLS scores but also the between-school variation predicted by school context variables.

The descriptive statistics provide strong evidence that there are large and

practically meaningful differences between schools not only in school contexts but also in first grade beginning- and end-of-year scores on dynamic indicators of early literacy skills. The 14 schools participating in this federally funded initiative, which was focused intensively on systematically improving reading instruction and outcomes for all students in Grades K-3, differed noticeably in the school context variables of interest, with a range of 49% in the percentage of incoming kindergarteners at risk and a range of 37% in the percentage of students eligible for free and reduced-price lunch, differences that are noteworthy for schools from a relatively homogeneous population of schools. Examination of the descriptives for the fall and spring of Grade 1 DIBELS scores indicates that the range of observed scores is larger in the spring, demonstrating that some schools may be more successful than others in supporting students' acquisition and mastery of the alphabetic principle.

The magnitude of variation between schools in first graders' performance on measures of alphabetic understanding and phonological recoding is commensurate with findings from previous school effects research (Teddlie et al., 2000). In particular, findings from the unconditional ANOVA models revealed that 9% of the variance observed in spring of Grade 1 NWF CLS scores was between schools, while 7.4% of the variance in WRWUC scores was also attributable to schools. Each student-level predictor added to the unconditional model independently was also a statistically significant predictor of end-of-year NWF scores that accounted for between 7% and 15.2% of the true between-school variance in those NWF scores. When all student-level predictors were added to each model, however, only allowing the effect of students' fall of Grade 1 NWF raw scores to vary randomly between students and the squared term of those scores

were significant predictors. The results from this model indicated that the relation between fall of Grade 1 NWF CLS and spring of grade 1 NWF CLS and WRWUC was not perfectly linear but was slightly curvilinear. The inclusion of these predictors in each model reduced the amount of within-school variation in spring of Grade 1 CLS by 42.8% and the amount of within-school variation associated with WRWUC by 35.8%.

Hierarchical Generalized Linear Modeling (HGLM) was used to address the second purpose of this study, investigating the possible absence measurement invariance associated with DIBELS NWF in the spring of Grade 1 by examining the difficulties of the first 20 pseudo-words on this benchmark probe. The results of the unconditional model revealed that the item difficulties, which ranged from -3.857 to -0.364, varied significantly. Moreover, the range of these item difficulties indicated that at the end of first grade, students with average levels of alphabetic understanding and phonological recoding skills had at least a 50% chance of accurately identifying all of the letter sounds that appeared in each of the first 20 nonsense words on the measure. Additionally, the item difficulty obtained for each pseudo-word was statistically significant ($p < 0.05$), meaning that the items were not equally difficult for all students. Results of the fully specified conditional model with all student-level predictors revealed that only two of the three student-level predictors—EL status and fall of Grade 1 NWF scores—were statistically significant predictors of students' item-level performance. More specifically, results indicated that EL students, on average, were predicted to respond correctly less often than their native English speaking peers and that for each correct letter sound read in the fall, a student increased his/her chance of responding correctly to an item in the spring by a value of 0.021 on the logit scale, compared to a student who read no correct

154

letter sounds. The implications of these findings, as well as hypotheses as to why the same student-level predictors were not statistically significant in the models predicting total outcome scores on NWF (i.e., CLS and WRWUC) and item-level performance, are presented in the Discussion section (Chapter V).

CHAPTER V

DISCUSSION

The purpose of this study was twofold: (a) to examine the degree to which the variance observed in first-grade students' performance on DIBELS NWF could be attributed to different school- and student-level predictors, and (b) to examine the possibility of measurement invariance by examining differences in the item difficulties generated for the first 20 pseudo-words on the spring of Grade 1 NWF benchmark probe. Data from 1,111 first-grade students enrolled in 14 elementary schools participating in the federally funded Oregon Reading First initiative were used to examine these research questions because these schools collected the data in such a way that allowed examination of item-level performance (see the Appendix for a copy of the probe used to collect these data). Answering these research questions required the use of two statistical modeling techniques: hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) to account for the nested nature of the data (i.e., students within schools), and hierarchical generalized linear modeling (HGLM; Kamata, 2001) to conduct multilevel Rasch analyses examining the effects of student-level predictors on item-level performance.

This chapter is divided into three sections. The first section includes a summary and interpretation of the findings obtained from the study, looking specifically at the contributions of the student- and school-level predictors. The second section addresses the limitations associated with the study, including threats to internal and external validity, and the third section includes a discussion of the implications of the findings and potential areas for future research.

156

Examining the Contributions of Student-Level Predictors

As noted in Chapter II, although research has indicated that the quality of classroom instruction is a "pivotal force" in the prevention of reading difficulties (Felton, 1993; National Research Council, 1998), individual student-level characteristics, such as initial skill level and language proficiency, are also likely to influence how quickly students become proficient readers. In particular, researchers have found that the level of knowledge and skill that a student brings to classroom instruction may be one of the strongest determinants of growth a student can be expected to make during the course of the school year (Good et al., 1998; Hedges & Hedberg, 2007). Several studies have found, for example, that a student's initial score on DIBELS NWF in the fall of first grade may account for anywhere between 26% and 58% of the variance observed in his or her scores on measures of Oral Reading Fluency in the spring of Grade 1 (Fien et al., 2008; Fien et al., 2010; Good et al., 2009).

Moreover, researchers (August & Shanahan, 2006; Gersten et al., 2005; Lesaux & Siegel, 2003) have also concluded that ELs can learn to read as well as their native English-speaking peers and that early reading measures may function similarly for both groups of students (Fien et al., 2008), providing evidence that ELs can become proficient readers in English and their progress can be monitored on commonly used measures of English literacy performance. Of particular interest to this study were the relations between students' EL status, fall of Grade 1 PSF score, and fall of Grade 1 NWF score and their overall performance on NWF in the spring of Grade 1 (as evidenced by CLS and WRWUC scores), as well as item-level performance.

Overall NWF Performance

With regard to students' overall performance on NWF, results of the sequential model-building process with CLS and WRWUC as outcome variables indicated that two of the three student-level predictors, EL status and fall of Grade 1 PSF raw score, were only statistically significant predictors of later student performance when entered individually. Adding EL status to the model with CLS as the outcome, for example, produced a model in which 7.84% of the variance remained between schools and accounted for 14.3% of the true between-school variance observed in spring of Grade 1 CLS scores. Similarly, EL status as a student-level predictor in the model with WRWUC as an outcome variable produced a model in which 6.8% of the variance remained between schools and accounted for 8.9% of the true, between-school variance observed in spring of Grade 1 WRWUC scores. When fall of Grade 1 PSF raw scores were entered in the model as the only student-level predictor for CLS, 8.4% of the observed variance was attributed to schools and 15.5% of the true, between-school variance in CLS was accounted for, and when entered as the only student-level predictor for WRWUC, these scores accounted for 7.0% of the true between-school variance (7.4% of the variance remained between schools). Each predictor entered independently, in other words, accounted for a meaningful portion of the variance observed in the outcomes of interest, but this significance dissipated when fully specified Level 1 models were constructed that included fall of Grade 1 NWF raw and quadratic scores; no longer were students' EL status and fall of Grade 1 PSF scores significant predictors for either CLS or WRWUC in the spring of Grade 1.

Item-Level NWF Performance

In contrast, two of the three student-level predictors of interest—EL status and fall of Grade 1 NWF scores—were statistically significant predictors of predicted item-level performance in the HGLM model. A statistically significant coefficient of -0.327 ($p < .05$) was associated with EL status in the conditional HGLM model, representing the prediction that EL students, on average, were estimated to have a lower score by -0.327 logits or, alternatively, that native English speakers were 1.39 times more likely to provide a correct response than EL students. A statistically significant coefficient of 0.021 ($p < 0.001$) was also associated with students' fall of Grade 1 NWF CLS scores, indicating that for each letter sound read correctly in the fall, a student increased his or her chance of getting an item correct in the spring by 0.021 on the logit scale compared to a student with a fall CLS score of 0. Alternatively, this means that a student who read the average number of letter sounds correctly in the fall, 77.16, is approximately 2.2 times more likely to get an item correct compared to a students whose fall CLS score was one standard deviation below the meaning (37.75) and 5.0 times more likely to do so than a student with a fall CLS score of 0. Similar to the HLM single-variable models with fall of Grade 1 PSF scores as a student-level predictor for later CLS and WRWUC scores, PSF was also not a statistically significant predictor of item-level performance ($\beta = 0.007$, $p = 0.057$).

Several explanations for these findings are possible. The non-significance of PSF as an initial predictor, for example, may be attributed to the reciprocal nature between phonological awareness and alphabetic understanding (Adams, 1990; Liberman et al.,

1989; Stanovich, 1986). This reciprocity is due primarily to the alphabetic structure of the English language that is represented by phonemes and graphemes; understanding that words are comprised of individual sounds is just as critical as the ability to link those sounds to letters (Perfetti, 1985). As Adams (1990) observed, "functional understanding of the alphabetic principle depends equally on knowledge of letters and on explicit awareness of phonemes because it depends integrally on the association *between* [emphasis added] them" (p. 304). With regard to the findings obtained in this study, it may be that students' knowledge and understanding of phonemes is adequately represented by their performance on NWF, for it is unlikely that students will perform well on NWF if they are not consciously aware that words—even pseudo-words—are made up of individual, discrete sounds (i.e., phonemic awareness). Researchers have obtained similar findings regarding the decreasing predictive ability of phonemic awareness of later reading skills (e.g., Oral Reading Fluency) when predictors more closely related to reading, such as Letter Sound Fluency and Nonsense Word Fluency, were considered (Perfetti, 1991; Powell-Smith & Cummings, 2007; Speece & Ritchey, 2005), although no studies appear to have examined this relation explicitly between PSF and NWF.

Explanations regarding the differential contribution of student EL status to overall performance on NWF as evidenced by CLS and WRWUC scores and item-level performance, however, may not be as straightforward. Two hypotheses may be plausible. First, previous research conducted examining student performance via CLS scores has revealed that NWF functions similarly for ELs and non-ELs (Fien et al., 2008). Fien et al. (2008), for example, found not only similar concurrent and predictive correlations

160

between NWF and other outcome measures of reading (i.e., ORF and reading comprehension) for ELs and non-ELs, but also that the amount of variance that performance on NWF accounted for on those outcome measures for the two groups of students was highly comparable. This may explain why student EL status was not a significant predictor of CLS and WRWUC performance. A second potential explanation related specifically to the statistical significance of EL status as a predictor for item-level performance is that ELs whose native language utilizes an alphabetic writing system, such as Spanish, may readily recognize the letter sounds that are common across the two languages, such as most consonants (Bialystok, Luk, & Kwan, 2005; Fien et al., 2008). Students' potential difficulty with correctly identifying English vowels, which is likely to contribute to the negative coefficient of -0.327 associated with their item-level performance because no vowel sounds are similar across both languages, may be masked in their CLS scores on NWF which considers their accuracy with consonant and vowel identification, especially when one considers that there are more CVC than VC words on the DIBELS NWF probe.

Although not an explicit focus of this study, the statistical significance of students' fall of Grade 1 NWF CLS score in all three models constructed for this study—with CLS and WRWUC as outcomes, as well as the HGLM model examining differences in item difficulties—speaks to the predictive utility of this measure for later performance on similar measures of reading. Researchers have reported, for example, predictive validity coefficients of .43 to .76 for various NWF probes administered at different points in time (Harn et al., 2008), as well as low to moderate coefficients with other indicators of reading performance (DMG, 2008). Correlations between fall of Grade 1 and spring of

161

Grade 1 NWF performance obtained in this study fall well within this range, with a correlation of $r = .63$ for CLS and $r = .57$ for WRWUC. Additionally, fall of Grade 1 NWF raw scores reduced the amount of within-school variance observed in CLS by 41.8% and the within-school variance observed in WRWUC by 33.8%; neither of these amounts is negligible and reveals that students' fall of Grade 1 scores are strongly related to their later performance in the spring of first grade. On the one hand, these results, particularly for CLS, are not surprising and speak to the predictive validity of NWF; because the measures and outcome scores are the same in the fall and spring, one would expect the fall score to have some explanatory power regarding the differences observed in spring scores. On the other hand, one might have expected fall CLS scores to explain more of the variance observed in WRWUC because firm knowledge of letter-sound correspondences is required to read the pseudo-words correctly. It may be, however, that this relationship was attenuated by the scoring procedure used for WRWUC in which students received full credit for one WRWUC even if they made an error blending the sounds on their first attempt but self-corrected and blended all sounds correctly within three seconds.

Examining the Contributions of School-Level Predictors: Schools Do Matter

As Reynolds et al. (2000) and other school effects researchers (Borman et al., 2003; Desimone, 2002) have noted, the idea that schools affect children's development is now a widespread assumption embraced by educational researchers and practitioners alike. What has yet to reach consensus, however, is not only how much of the variance in students' performance can be attributed to schools but also which school-level variables

162

need to be considered in examining the influence of schools on children's development. Researchers have found, for example, that anywhere from 8% to 36% of the variance in student achievement can be attributed to schools (Fien et al., 2009; Teddlie et al., 2000), depending on the grades examined and the outcome(s) of interest. Two school-level variables were considered in this study: (a) the percentage of students eligible for free and reduced-price lunch as an indicator of socioeconomic status, and (b) the percentage of incoming kindergarten students at risk for later reading difficulties as an indicator of initial student achievement.

Contributions of Percent FRL

Although research has indicated that most school effects research attempts to account for context differences by including context variables such as the socioeconomic composition of the student population and the differential effectiveness of schools by controlling for prior attainment or initial status (Teddlie et al., 2000), the effects of these school-level variables remain unclear. Some studies have indicated, for example, that a school's socioeconomic status (as measured by percentage of students eligible for free and reduced lunch) is not significantly related to between-school variation in student outcomes (Borman et al., 2003; Fien et al., 2009), while others have found the opposite (Willms & Raudenbush, 1989). Findings obtained from this study, for example, revealed that including Percent FRL as a fixed predictor on the intercept and NWF slope explained 75% of the 9% between-school variance observed in spring of Grade 1 CLS scores and 68.6% of the 7.4% between-school variance observed in spring of Grade 1 WRWUC scores.

Beyond the overall variation attributed to Percent FRL, the practical implications associated with Percent FRL presented in Tables 10 and 16 warrant further discussion. These results indicate, for example, that students attending schools with greater percentages of students eligible for free and reduced-price lunch are predicted to have lower CLS and WRWUC scores in the spring of Grade 1 compared to their similarly performing peers attending schools with less challenging contexts, results that are consistent with previous research (Juel, 1988; Stanovich, 1986). More interesting, however, is the interaction effect observed between Percent FRL and students' fall of Grade 1 NWF CLS scores in the prediction of spring of Grade 1 CLS and WRWUC scores. The fact that the coefficient for Percent FRL on the intercept is not statistically significant means that there is no correlation between predicted spring NWF scores and Percent FRL for students' whose fall of Grade 1 NWF score was 0. For these students (and presumably students' with low fall of Grade 1 NWF scores), their spring NWF scores are predicted to be relatively consistent regardless of Percent FRL in their school.

Additionally, the negative, statistically significant correlation of Percent FRL on NWF slope means that as fall NWF scores increase, the correlation between predicted spring NWF scores and Percent FRL becomes more negative. Students with higher fall of Grade 1 NWF scores, in other words, are predicted to have lower spring NWF CLS and WRWUC scores if they attend schools with above average percentages of FRL, whereas students with lower fall of Grade 1 NWF scores appear to be relatively less influenced by the Percent FRL in their school. As presented in Chapter IV, one potential explanation for this finding may be that students' attending schools with lower percentages of students eligible for free and reduced price lunch have, on average, higher fall CLS scores and

164

therefore have less room to grow during the school year. A second potential explanation for this finding may be related to the allocation of instructional resources (e.g., personnel, materials, time). If schools have designed their first-grade systems to provide the most intensive supports for students who have demonstrated the greatest risk in the fall of Grade 1, students with higher fall of Grade 1 NWF scores who may appear to be less at risk for future reading difficulties may not receive the instructional supports necessary to maintain a positive trajectory in performance. This finding may, however, be less important for students who have met or exceeded the benchmark goal of 50 CLS in the fall, for research has indicated that once students have achieved a sufficient level of automaticity with decoding tasks (such as letter-sound correspondence knowledge and phonological recoding) subsequent gains may not be meaningful (Fien et al., 2010).

Contributions of Percent K Risk

The findings associated with the percentage of incoming kindergarten students at risk as a school-level predictor in the models with CLS and WRWUC may, at first glance, be counterintuitive but definitely deserve further discussion. One might expect, for example, that students attending schools with higher percentages of incoming at-risk kindergarten students would be predicted to have lower spring of Grade 1 NWF scores because of the sheer number of students likely needing explicit, systematic, and intensive instructional support to even approach grade-level proficiency. Findings from this study indicate, however, that the opposite is true. Students in this sample appear to have benefited from attending schools with higher percentages of incoming at-risk kindergarten students. That is to say, regardless of school socioeconomic context,

165

students are predicted to obtain higher CLS and WRWUC scores at the end of Grade 1 if they attend schools with a greater than average percentage of at-risk kindergarten students.

There are three potential explanations for this unexpected finding. First, schools with greater percentages of incoming students at risk may have established a strong, healthy *system* to support these low-performing students via the provision of increased amounts of explicit, systematic, intensive instruction, frequent collection of student reading data to monitor student progress, and/or the targeted allocation of experienced and well-trained personnel toward the lower grades (i.e., Grades K and 1). Second, as schools had been collecting benchmark and progress monitoring data for students in Grades K-3, it may have become evident that students in the upper primary grades (i.e., Grades 2 and 3) were not making expected rates of growth and a school-wide decision was made to follow the recommendations of researchers (Foorman, Breier, & Fletcher, 2003; Foorman et al., 1997) by providing targeted, instructional supports *early* in an effort to prevent later reading difficulties. Third, as a result of observing a "slump" in second- and third-grade DIBELS scores during the 2006-2007 and 2007-2008 school years, the ORFC established two project-level goals in the fall of 2008 that may have influenced the quality, intensity, and/or focus of instruction provided to students in Grades K and 1: (a) increase student automaticity in whole word reading in kindergarten; and (b) provide targeted instruction, based on student need, to all strategic and intensive first-grade students (Travers, 2009). A majority of the on-site and remote technical assistance provided to participating schools during the 2008-2009 and 2009-2010 school years was related to these project-level goals with an explicit focus on increasing the

166

performance of students in the lower primary grades as a means for establishing a solid foundation for success in the later grades.

Regardless of the impetus for this improvement, students of all skill levels attending schools with higher percentages of incoming kindergarten students at risk are predicted to earn higher NWF CLS and WRWUC scores in the spring of first grade than their peers in schools with less than average percentages of at-risk kindergarten students. This finding is important in and of itself, but also has implications for lower performing students in schools with less than average percentages of at-risk kindergarten students, for these students may actually fare *worse* than they would if they were to attend a school with a higher percentage of at-risk kindergarten students because the instructional supports they need to succeed may not be firmly established (i.e., resources may be allocated to a different population of students). Students may, in other words, benefit from attending schools with similar populations of students because there is a stronger possibility that instructional resources and supports (i.e., materials, personnel, instructional time, etc.) will be available to support their needs.

Examining Item Variability: Not All Items Are Created Equal

The primary objective of Research Question 3 was to examine whether the first 20 pseudo-words on the spring of Grade 1 NWF benchmark probe were of comparable difficulty, for not only should students with equal levels of skill, knowledge, or ability on the construct of interest earn the same test score but also have equal probabilities of responding to items correctly, irrespective of group membership (AERA et al., 1999). Examination of the item difficulties obtained from the unconditional model with no

student-level predictors revealed a substantial range in difficulties (-3.857 to -0.364), all of which were statistically significant at $p < 0.001$. These results revealed significant variability in item difficulties for the first 20 pseudo-words, variation that was also evident for some items when student EL status and fall of Grade 1 NWF raw scores were added to the model as student-level covariates. In particular, results of this model reveal the likelihood that EL students will respond correctly less frequently than their native English speaking peers and that students' performance in the fall is a significant predictor of their item-level performance in the spring of grade 1. Moreover, results of the HGLM indicate that the inclusion of these student-level predictors to the model reduced the variance between schools to a non-significant level and explained the differences observed in item difficulties.

Limitations of the Study

Several limitations associated with this study not only pose threats to internal and external validity but also influence the overall interpretability of the findings. One such limitation is the relatively small number of schools (14), which may not only affect issues of power, but also may limit the generalizability of the findings, as discussed later in this section. A second, and perhaps more notable limitation was the fact that the models employed in this study did not account for the full hierarchical nature of the data in which students were nested within classrooms/reading groups within schools. Although this is undoubtedly an important contextual variable that warrants further investigation, especially considering that providing instruction in small groups can have positive effects on student achievement, particularly for students at risk (Elbaum, Vaughn, Hughes, &

168

Moody, 1999; Foorman & Torgesen, 2001; Vaughn et al., 2003), this level of examination was not possible due to the flexible, homogeneous, cross-class grouping used in participating schools to provide students with differentiated instruction to meet their individual needs.

Threats to Internal Validity

Two potential threats to internal validity associated with this study are testing (or practice effects), and the challenges posed by fluency-based outcome measures.

Testing and Practice Effects

One potential threat to internal validity is that of testing, or practice effects that may result when students take a test multiple times or simply become better at something the more often they do it (Shadish, Cook, & Campbell, 2002). Although this is definitely a possibility in this study because all students were administered the DIBELS NWF assessment in the fall, winter, and spring of first grade and some were progress monitored on a more frequent basis, one way that this threat is controlled is by the use of alternate forms. More specifically, each NWF benchmark probe contains, for the most part, a different set of words that utilize the same letters of the alphabet representing the most common sounds of the English language. For those students who were progress monitored on NWF between benchmark data-collection time points, 20 alternate forms of NWF were available for use.

Fluency-Based Outcome Measures

Although not characterized as a common threat to internal validity (Shadish, Cook, & Campbell, 2002) an additional threat that needs to be considered in the context of this study is the fluency-based nature of the outcome, DIBELS NWF. As stated earlier, DIBELS NWF is a standardized, timed assessment during which students have 1 minute to read as many pseudo-words as they can as accurately as possible. Because it is a timed test, the number of words that a student has the opportunity to read depends heavily on their skill level; students with greater mastery of and automaticity with letter-sound correspondences and decoding skills are likely to be exposed to more pseudo-words than students with lower levels of skill. This issue is of particular importance for this study as it affects the degree to which the difficulty of the items (or pseudo-words) can be appropriately measured; some items may appear more difficult for others based solely on the fact that fewer students reached those items during the 1-minute timing. The effort was made in this study to address the fluency-based nature by estimating the difficulties of the first 20 pseudo-words, which 70.5% of students in the sample read, because that is equal to the 50 CLS a student would need to read correctly to reach the DIBELS benchmark goal for this measure, however, alternative methods for determining how to account for issues related to fluency, speededness, and automaticity should be pursued.

Threats to External Validity

The greatest threat to external validity associated with this study is generalizability, and on two different levels: generalizability of performance on different

170

outcomes and generalizability to different settings.

Generalizability of Student Performance

This threat of generalizability associated with student performance is due to the fact that, for this study, students' acquisition of the alphabetic principle is examined only in the context of the DIBELS NWF measure. It is possible, for example, that a student's performance would not generalize to other measures of pseudo-word reading, such as the Test of Word Reading Efficiency (TOWRE) phonetic decoding subtest or the Woodcock Reading Mastery Test-Revised (WRMT-R) word attack subtest. Additionally, although research has indicated that measures of pseudo-word reading are among the most common for examining students' decoding skills (NICHD, 2000; Siegel, 1998; Vellutino, 1991), some researchers (Fuchs, Fuchs, & Compton, 2004) have argued that nonsense word fluency is not a useful task for examining concurrent and predictive validity on later measures of reading performance and that measures of real word reading ought to be used instead.

Generalizability to Other Settings

As noted earlier, this study was conducted using data from schools participating in the Oregon Reading First initiative that received on-site technical assistance from Oregon Reading First Center staff during the 2008-2009 and 2009-2010 school years. More specifically, much of the technical assistance these schools received was focused on helping them interpret students' performance to inform instructional grouping decisions; these schools received specially designed reports that provided detailed

information for each student about the various word reading strategies students used to read the pseudo-words to facilitate this process.

It is also worth noting that these schools participating in the Oregon Reading First initiative had received extensive training in providing highly specified, research-based instruction to all students and how to differentiate this instruction based on student need. As Fien et al. (2008) noted, this instruction included a strong and explicit emphasis on phonological awareness and decoding skills, both of which are captured in students' performance on NWF. It is possible that these results may not generalize to other school settings that (a) did not provide explicit, systematic, research-based instruction; (b) did not provide teachers with extensive professional development focused on phonological awareness and alphabetic understanding; and/or (c) did not use NWF data to screen students, monitor progress, and adjust instruction to meet student needs.

Implications

Findings from this study indicate that student-level characteristics, school-level characteristics, and the structure of the measurement used to assess students' alphabetic understanding and phonological recoding knowledge and skills contribute to the amount of variance observed in student performance, both between and within schools. Results indicate, for example, that students' fall of Grade 1 NWF performance is the most significant student-level predictor of spring of Grade 1 NWF performance (for both CLS and WRWUC) and that all pseudo-words that appear on a NWF probe may not be of comparable difficulty for all students. These findings lend support to a closer examination of student performance on DIBELS NWF by teachers and school leaders,

172

particularly for students who have demonstrated difficulty learning to read. Item-level examination of student performance on DIBELS NWF may provide educators with some insights regarding which letter-sound correspondences students are struggling with, information that can then be used to provide targeted, explicit, instruction to fill in the gaps in students' alphabetic principle knowledge.

Furthermore, results indicate that schools matter, as evidenced by the complex and significant relationships observed between the two school-level predictors in this study—percentage of incoming kindergarten students and percentage of students eligible for free and reduced-price lunch—and students' performance on NWF in the spring of first grade. This finding is particularly notable considering the 14 schools that participated in this study came from a relatively homogeneous population of schools that had been required to meet stringent poverty and student performance criteria. Although each of these variables accounted for noticeable portions of the variance observed in NWF CLS and WRWUC scores in the spring of Grade 1, further investigation into additional context variables associated with schools is also needed. It is possible, for example, that any of the following context variables may account for a portion of the remaining unexplained variance in the final models for NWF CLS and WRWUC: (a) amount of instructional time allocated to reading instruction, and phonics instruction in particular; (b) reading program(s) used to provide instruction; (c) the amount of on-site technical assistance schools received from ORFC regional coordinators; and (d) the distribution and allocation of resources (e.g., whether more personnel were available to teach smaller groups of students in the lower grades). Although additional information would be useful, these findings lend support to arguments made about the importance of

schools and their influence on student achievement.

Directions for Future Research

Although these findings support the position that schools, students, and measures *do* matter and contribute to the variance observed in students' performance, it is clear that additional research is warranted to more comprehensively address the research questions posed in this study. In particular, additional research in the following areas would be valuable: (a) the instructional context of schools, (b) further examination of student performance on the spring of Grade 1 NWF probe to examine the item difficulties of additional pseudo-words, (c) examination of item-difficulties for only those pseudo-words students read as whole words, and (d) consideration of other potential student-level characteristics that may be influencing student performance on DIBELS NWF.

As noted during the discussion about potential implications associated with this study, it would be worthwhile to investigate further the context within which instruction was provided to determine what other factors may account for the variance observed in student performance. Research, has indicated, for example, that reading programs may not attend to and incorporate all research-based recommendations and findings regarding content and instructional design principles (Smith et al., 2001) and that significant differences in student performance on early reading measures may be explained, in part, by the program(s) within which they have received reading instruction (Crowe, Connor, & Petscher, 2009). As noted in Chapter III, schools participating in the Oregon Reading First initiative were required to adopt core, supplemental, and intervention materials from an approved list of programs but were given liberty to decide which programs to adopt

174

and how they would implement them. Additional instructional context variables that may warrant further investigation include (a) amount of instructional time allocated to reading instruction; (b) instructional grouping practices; (c) teacher fidelity of reading program implementation; (d) the quantity and quality of professional development teachers received regarding program implementation and phonics instruction; and/or (e) whether teachers collected and *used* various sources of student performance data (i.e., DIBELS progress-monitoring assessments, in-program assessments, etc.) to inform instruction.

In addition, because examination of the first 20 pseudo-words on the spring of Grade 1 NWF benchmark probe revealed significant variation in item difficulties, further examination of the NWF measures is also warranted. In particular, it would be worthwhile to utilize alternative methods to control for the fluency-based nature of the outcome measure to examine the item difficulties of pseudo-words beyond the first 20. Moreover, similar analyses could be conducted with other DIBELS NWF probes (benchmark and progress-monitoring probes are available for Grades K, 1, and 2) to determine if variability in item difficulties is associated with those measures as well. It may also be informative to conduct similar analyses with the recently released DIBELS Next (Good & Kaminski, 2011) NWF probes to determine if the efforts made to account for differences in item difficulty via the random stratification of pseudo-words has addressed issues of measurement variance. Lastly, because differences in performance were predicted for ELs compared to native English speakers, it would also be worthwhile to conduct invariance testing of item difficulties to determine if there are systematic differences in item difficulties based on group membership.

Lastly, this study touched briefly upon the notion that words starting with stop

sounds (e.g., /p/, /b/, /k/, /g/, /d/, and /t/) are hypothesized to be more difficult for students than words starting with continuous sounds (e.g., /m/, /n/, /s/, /f/, /r/, etc.; Allor & McCathren, 2003; Smith et al., 1998). Preliminary examination of the item difficulties obtained in this study did not readily support this hypothesis. To examine this hypothesis more appropriately, however, requires the estimation of item difficulties for pseudo-words that students read as whole words (i.e., did not attempt at the sound level) because the proposed differences in difficulty are thought to occur when students are learning to *blend* sounds, or phonologically recode.

Conclusions

This study examined the student- and school-level predictors of students' performance on DIBELS NWF in the spring of Grade 1 at both the sound (total Correct Letter Sounds) and word (Words Read as Whole Units Correctly) levels. Also of interest was possible variation in item-level performance, examined via comparisons of item difficulties of the first 20 pseudo-words. To examine the research questions posed, I analyzed data from 1,111 first-grade students enrolled in 14 elementary schools participating in the federally funded Oregon Reading First initiative and implementing a comprehensive, preventive, and systemic approach to research-based reading instruction.

Four general conclusions emerge from this study. First, schools are an appropriate unit of analysis and account for meaningful portions of variance observed both *between* and *within* schools in students' performance on DIBELS NWF in the spring of first grade. Second, complex relationships exist between school-level context variables and student performance, including higher predicted NWF scores for students attending

176

schools with greater than average percentages of incoming kindergarten students and a complex interaction between student NWF performance and school socioeconomic status. Third, student's fall of Grade 1 performance on NWF accounted for the largest portion of variance observed in their spring of Grade 1 NWF CLS and WRWUC scores compared to other student-level characteristics, which speaks to the predictive validity of the measure. Fourth, significant variation was observed in item-level performance on the first 20 pseudo-words, although this variance was reduced to a non-significant level with the inclusion of student EL status and fall of grade 1 CLS scores as student-level covariates, providing a rationale for similar examination of additional pseudo-words and NWF probes.

APPENDIX

SAMPLE OF NWF PROBE USED TO COLLECT STUDENT DATA

DIBELS Nonsense Word Fluency
First Grade
Winter Benchmark

Grade

Directions:
* Please score with dashes and slashes per DIBELS guidelines.

Examiner:
Date: ___ / ___ / ___

f o j	h o n	t u m	l e n	a f	___/14
s u v	k a m	r e s	k i c	f a v	___/15
i d	w o d	n e j	s o k	w i f	___/14
b i s	w e m	a v	j a j	a c	___/13
l u j	w a n	u c	b a s	i d	___/13
b e f	z i j	v e n	w i l	v e p	___/15
l o d	o s	h i c	v o s	s o v	___/14
r i l	v e s	j o s	s e j	j e b	___/15
p a c	e c	d o k	v u t	f e g	___/14
v i b	k o v	h u v	s e b	j i t	___/15

Total Correct Letter Sounds (CLS): ▢▢▢

Total words recoded completely and correctly (WRC): ▢▢▢

If student reads all words before 1 minute has ended, please continued on other side of page

REFERENCES CITED

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: Massachusetts Institute of Technology.

Alliance for Excellent Education. (2009). *Oregon high schools.* Retrieved from http://www.all4ed.org/about_the_crisis/schools/state_information/Oregon

Allor, J. H., & McCathren, R. B. (2003). Developing emergent literacy skills through storybook reading. *Intervention in School and Clinic, 39*, 72–79.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Federation of Teachers. (1998). *Building on the best, learning from what works: Seven promising reading and English language arts programs.* Washington, DC: American Federation of Teachers.

Ardoin, S. P., Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38,* 266-283.

August, D., & Shanahan, T. (2006). Introduction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 1–42). Mahwah, NJ: Lawrence Erlbaum.

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

Baker, S. K., Smolkowski, K., Smith, J. M., Fien, H., Kame'enui, E. J. & Thomas Beck, C. (in press). The impact of Oregon Reading First on student reading outcomes. *Elementary School Journal.*

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26,* 49–66.

Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of readability statistics and equating for alternate forms. *Journal of School Psychology, 47,* 1-17. doi:10.1016/j.jsp.2008.09.001

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading, 9,* 43-61.

Billings, D. (2002). Oregon Reading First grant application. U.S. Department of Education Reading First Program No. 84.357.

Blachman, B. A., Tangel, D. M, Ball, E.W., Black, R., & McGraw, C. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal, 11,* 239–273.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 125–230.

Burke, M., & Hagan-Burke, S. (2007). Concurrent validity for first grade early literacy indicators. *Assessment for Effective Intervention, 32,* 66-77.

Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*, 403–414.

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81,* 313–321.

Cardoso-Martins, C. (2001). The reading abilities of beginning readers of Brazilian Portuguese: Implications for a theory of reading acquisition. *Scientific Studies of Reading, 5*, 289–317. doi: 10.1201/S1532799XSSR0504_01

Carnine, D. W., Silbert, J., Kame'enui, E. J., Tarver, S. J., & Jungjohann, K. (2006). *Teaching struggling and at-risk readers: A direct instruction approach.* Upper Saddle River, NJ: Pearson.

Chard, D. J., & Harn, B. A. (2008). Project CIRCUITS: Center for Improving Reading Comprehensive Using Intensive Treatments Schoolwide. In C. R. Greenwood, T. Kratchowill, & M. Clements (Eds.) *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 143-172). New York, NY: Guilford.

Chard, D. J., & Osborn, J. (1999). Phonics and word recognition in early reading programs: Guidelines for accessibility. *Learning Disabilities Research and Practice, 14*, 107–117.

Chard, D. J., Simmons, D. C., & Kame'enui, E. J. (1998). Word recognition: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells*

us about children with diverse learning needs: Bases and basics (pp. 141–168). Mahwah, NJ: Lawrence Erlbaum.

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,* 174–188.

Clements, M., & Kratchowill, T. (2008). Multi-tier prevention models: Implications and future perspectives. In C. R. Greenwood, T. Kratchowill, & M. Clements (Eds.) *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 269-290). New York, NY: Guilford.

Clifford, M. M. (1990). Students need challenge, not easy success. *Educational Leadership, 48*, 22–26.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). Equality of educational opportunity. Washington, DC: National Center for Educational Statistics.

Comprehensive School Reform Quality Center. (2006a). *CSRQ Center report on elementary school comprehensive school reform models.* Washington, DC: Comprehensive School Reform Quality Center & American Institutes for Research.

Comprehensive School Reform Quality Center. (2006b). *Executive summary: CSRQ Center report on middle and high school comprehensive school reform models.* Washington, DC: Comprehensive School Reform Quality Center & American Institutes for Research.

Coyne, M., Kame'enui, E. J., & Simmons, D. C. (2001). Prevention and intervention in beginning reading: Two complex systems. *Learning Disabilities Research and Practice, 16*, 62–73.

Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2004). Improving beginning reading instruction and intervention for students with LD: Reconciling "all" with "each". *Journal of Learning Disabilities, 37,* 231-239.

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*, 90–104. doi: 10.11777/00222194040370020101

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Crowe, E. C, Connor, C. M, & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47,* 187–214. doi: 10.1016/j.jsp.2009.02.002

Cummings, K., Dewey, E., Latimer, R., & Good, R. H. (in press). Pathways to word reading and decoding: The roles of automaticity and accuracy. *School Psychology Review.*

Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology, 50,* 429–444.

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.

Curtis, M. E. (1990). Development of components of reading skill. *Journal of Educational Psychology, 72*, 656–669.

D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievement. *School Effectiveness and School Improvement, 11,* 197-235.

Defior, S., & Tudela, P. (1994). Effect of phonological training on reading and writing acquisition. *Reading and Writing: An Interdisciplinary Journal, 6,* 299–320.

Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities, 39*, 447–466.

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72,* 433–479.

Dressler, C., & Kamil, M. (2006). First- and second-language literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second language learners: Report of the National Reading Panel on language-minority children and youth* (pp. 197–237). Mahwah, NJ: Lawrence Erlbaum.

Dynamic Measurement Group. (2008). *DIBELS 6th edition technical adequacy information* (Tech. Rep. No. 6). Eugene, OR: Author. Retrieved from http;//dibels.org/pubs.html

Education Commission of the States. (2002). *State literacy programs: Oregon Reading Initiative—Reading Together!* Retrieved from http://www.ecs.org/dbsearches/ SearchInfo/Literacy_ProgramProfile.asp?ProgID=44

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale, NJ: Lawrence Erlbaum.

Ehri, L. C. (1997). Sight word learning in normal readers and dyslexics. In. B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Applications for early reading intervention* (pp. 163–190). Mahwah, NJ: Lawrence Erlbaum.

Ehri, L. C. (1999). Phases of development in learning to read words. In J. Oakhill & R. Beard (Eds.), *Reading development and the teaching of reading: A psychological perspective* (pp. 79–108). Oxford, England: Blackwell.

Ehri, L. C. (2005a). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Malden, MA: Blackwell.

Ehri, L. C. (2005b). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*, 167–188.

Ehri, L. C., & McCormick, S. (1998) Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly, 14*, 135–163.

Ehri, L. C., & Snowling, M. J. (2005). In C. Addison Stone, E. R. Silliman, B. J. Ehren, & K. Appel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 433–460). New York, NY: Guilford.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children, 65,* 399-415.

Embretson, S. E., & Riese, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Felton, R. H. (1993). Effects of instruction on decoding skills of children with phonological-processing problems. *Journal of Learning Disabilities, 26*, 583–589.

Fielding-Barnsley, R. (1997). Explicit instruction in decoding benefits children high in phonemic awareness and alphabet knowledge. *Scientific Studies of Reading, 1*(1), 85–98.

Fien, H., Baker, S. K., Smolkowski, K., Mercier Smith, J. L., Kame'enui, E. J., & Thomas Beck, C. (2008). Using Nonsense Word Fluency to predict reading proficiency in kindergarten through second grade for English learners and native English speakers. *School Psychology Review, 37*, 391–408.

Fien, H., Kame'enui, E. J., & Good, R. (2009). Schools engaged in school-wide reading reform: An examination of the school and individual student predictors of kindergarten early reading outcomes. *School Effectiveness and School Improvement, 20,* 1-25.

Fien, H., Park, Y., Baker, S. K., Smith, J. L., Stoolmiller, M., & Kame'enui, E. J. (2010). An examination of the relation of Nonsense Word Fluency initial skill status and growth to reading outcomes for beginning readers. *School Psychology Review, 39,* 631–653.

Finegan, E. (2008). *Language: Its structure and use* (5th ed.) Boston, MA: Thomson Wadsworth.

Fitzgerald, J. (1993). Views on bilingualism in the United States: A selective historical review. *Bilingual Research Journal, 17,* 35–56.

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*, 613–639. doi: 10.1207/S15326942DN242&3_06

Foorman, B. R., Chen, D., Carlson, C., Moats, L., Francis, D. J., & Fletcher, J. M. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing: An Interdisciplinary Journal, 16*, 289–324.

Foorman, B. R., Francis, D. J., Fletcher, J. M, Schatschneider, C, & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37–55.

Foorman, B. R., Francis, D. J., Novy, D. M, & Liberman, D. (1991). How letter-instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology, 83*, 456–469.

Foorman, B. R., Francis, D. J., Shaywitz, S. E., Shaywitz, B. A., & Fletcher, J. M. (1997). The case for early reading intervention. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early reading intervention* (pp. 243–264). Mahwah, NJ: Lawrence Erlbaum.

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in Grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31,* 1–29.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice, 16,* 203–212.

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,* 315-342.

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). Malden, MA: Blackwell.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27,* 216-228.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & German, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27–48.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study: Interim Report.* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute for Education Sciences, U.S. Department of Education.

Genessee, F., Geva, E., Dressler, C., & Kamil, M. (2006). Synthesis: Cross-linguistic relationships. In D. August & T. Shanahan (Eds.), *Developing literacy in second language learners: Report of the National Reading Panel on language-minority children and youth* (pp. 153–174). Mahwah, NJ: Lawrence Erlbaum.

Gersten, R., Baker, S. K., Haager, D., & Graves, A. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners: An observational study. *Remedial and Special Education, 26*, 197–206.

Good, R. H., Baker, S. K., & Peyton, J. A. (2009). Making sense of Nonsense Word Fluency: Determining adequate progress in early first-grade reading. *Reading and Writing Quarterly, 25*, 33–56. doi: 10.1080/10583560802491224

Good, R. H. III, & Kaminski, R. A. (Eds.) (2002). *Dynamic indicators of basic early literacy skills (*6th ed.) *Administration and scoring guide*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu

Good, R., H. III, & Kaminski, R. A. (Eds.). (2007). *Dynamic indicators of basic early literacy skills* (6th ed. rev.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu

Good, R. H. III, & Kaminski, R. A. (2011). *DIBELS Next assessment manual*. Eugene, OR: Dynamic Measurement Group. Retrieved from http:/www.dibels.org

Good, R. H. III, Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *Educational and Child Psychology, 15*, 56–70.

Greenwood, C. R., Horner, R. H., & Kratchowill, T. (2008). Introduction. In C. R. Greenwood, T. Kratchowill, & M. Clements (Eds.) *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 3-30). New York, NY: Guilford.

Grim illiteracy statistics indicate Americans have a reading problem. (2007). Retrieved from http://education-portal/articles/Grim_Illiteracy_Statistics_Indicate_Americans_Have_a_ Reading_Problem.html

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and Non-Hispanic Students. *Journal of Special Education, 39*, 66–85.

Hagan-Burke, S., Burke, M., & Crowder, C. (2006). The convergent validity of the Dynamic Indicators of Basic Early Literacy Skills and the Test of Word Reading Efficiency for the beginning of first grade. *Assessment for Effective Intervention, 31,* 1-15.

Harn, B. A., Stoolmiller, M., & Chard, D. C. (2008). 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

Hedges, L., & Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials in educational research. *Educational Evaluation and Policy Analysis, 29,* 60-87.

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.

Juel, C. (1991). Beginning reading. In R. Barr, M. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 759–788). Mahwah, NJ: Lawrence Erlbaum.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement, 38,* 79–93.

Kame'enui, E. J., & Simmons, D. C. (1990). *Designing instructional strategies: The prevention of academic learning problems.* Englewood Cliffs, NJ: Macmillan.

Kame'enui, E. J., Simmons, D. C., & Coyne, M. D. (2000). Schools as host environments: Toward a school-wide reading improvement model. *Annals of Dyslexia, 50,* 31–52.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Laugle, K. M. (2009). *Measuring the alphabetic principle: Mapping behaviors onto theory* (Unpublished doctoral dissertation). University of Oregon, Eugene.

Learning First Alliance. (1998, Spring/Summer). Every child reading: An action plan of the Learning First Alliance, *American Educator,* 52–63.

Lesaux, N., Koda, K., Siegel, L., & Shanahan, T. (2002). Development of literacy. In D. August & T. Shanahan (Eds.) *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 75–122). Mahwah, NJ: Lawrence Erlbaum.

Lesaux, N., & Siegel, L. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39,* 1005–1019.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp.1–33). Ann Arbor, MI: University of Michigan Press.

Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. *Reading Research Quarterly, 26*, 234–250.

Literacy Leadership State Steering Committee & Oregon Department of Education. (2009). *Oregon K-12 Literacy Framework.* Salem, OR: Oregon Department of Education.

Logan, G. D. (1997). Automaticity in reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly, 13*, 123–147.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263–284.

Lyon, G. R. (2009). *The NICHD research program in reading development, reading disorders, and reading instruction*. Washington, D.C.: National Center for Learning Disabilities. Retrieved from http://www.ncld.org/ld-basics/ld-aamp-language/reading

McCandliss, B., Beck, I. L., Sandak, R., & Perfetti, C. (203). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading, 7*, 75–104.

Múthen, L. K., & Múthen, B. O. (2010). Mplus: Statistical analysis with latent variables: User's guide. Los Angeles, CA: Múthen & Múthen.

National Center for Educational Statistics. (2003). 2003 National Assessment of Adult Literacy. Retrieved from http://nces.ed.gov/naal/kf_demographics.asp

National Center for Educational Statistics. (2010a). *Digest of education statistics 2009* (NCES 2010-0103). Retrieved from http://nces.ed.gov/programs/digest/d09/ch_2.asp

National Center for Educational Statistics. (2010b). The *nation's report card reading 2009: National assessment of educational progress at Grades 4 and 8.* Washington, DC: U.S. Department of Education, Institute on Educational Sciences.

National Center for Educational Statistics (2010c). *NAEP data explorer* [Data file]. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/report.aspx.

National Center for Educational Statistics (2010d). Common Core of Data Public Schools. Retrieved October 15, 2010 from http://nces.ed.gov/ccd/schoolsearch/.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office. Retrieved from http://www.nichd.nih.gov/publications/nrp/report.htm

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

No Child Left Behind Act (NCLB) Act of 2001, Pub. L. No. 107-110, §115, Stat. 1425 (2002).

O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology, 87,* 202–217.

Oregon Department of Education. (2009). *Statewide report card 2008-2009: An annual report to the legislature on Oregon public schools.* Retrieved from the Oregon Department of Education website: http://www.ode.state.or.us/data/annreportcard/rptcard2009.pdf

O'Shaughnessy, T. E., & Swanson, H. L. (2003). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities, 33*, 257–277.

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

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

Perfetti, C. A. (1991). Representations and awareness in the acquisition of reading competence. In L. Rieben & C. A. Perfetti (Eds.) *Learning to read: Basic research and its implications* (pp. 33-46). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In. C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford, England: Oxford University Press.

Petscher, Y., & Kim, Y. (2010). The utility and accuracy of oral reading fluency score type in predicting reading comprehension. *Journal of School Psychology, 49,* 107-129.

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

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23,* 326-338.

Powell-Smith, K. A., & Cummings, K. D. (2007, February). What's PSF got to do with it? A look at the contribution of DIBELS Phoneme Segmentation Fluency to first grade reading outcomes. Poster presented at the 15th Annual Pacific Coast Research Conference, San Diego, CA.

Raudenbush, S. W., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong,Y., & Congdon, R. T. (2004). HLM 6.08: Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20,* 307-335.

Reading Excellence Act of 1998 (P. L. 105-277). *Title VIII, Part C. Amendment to ESEA for reading and literacy grants.*

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An introduction to school effectiveness research. In C. Teddlie & D. Reynolds (Eds.) *The international handbook of school effectiveness research* (pp, 3-25). New York, NY: Falmer.

Ritchey, K. D. (2008). Assessing letter sound knowledge: A comparison of letter sound fluency and nonsense word fluency. *Exceptional Children, 74*, 487–506.

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/explicit teaching. *Reading and Writing Quarterly, 25*, 125–138.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly, 13*, 107–122.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48*, 115–136.

Seymour, P. H. K. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.) *The science of reading: A handbook* (pp. 296–315). Malden, MA: Blackwell.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Siegel, L. S. (1998). Phonological processing deficits and reading disabilities. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 141–160). Mahwah, NJ: Lawrence Erlbaum.

Simmons, D. C., & Kame'enui, E. J. (1998). Introduction. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics.* (pp. 1–17). Mahwah, NJ: Lawrence Erlbaum.

Simmons, D. C., Kame'enui, E. J., Good, R. H. III, Harn, B. A., Cole, C., & Braun, D. (2002). Building, implementing, and sustaining a beginning improvement model: Lessons learned school by school. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems: II. Preventive and remedial approaches* (pp. 537–550). Washington, DC: National Association of School Psychologists.

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Edwards Santoro, L., . . . Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities, 40*, 331–347.

Simmons, D. C., Kuykendall, K., King, K., Cornachione, C., & Kame'enui, E. J. (2000). Implementation of the schoolwide reading improvement model: "No one ever told us it would be this hard!" *Learning Disabilities Research and Practice, 15,* 92-100. doi: 10.1207/SLDRP1502_5

Smith, S. B., Simmons, D. C., Gleason, M. M., Kame'enui, E. J., Baker, S. K., Sprick, M., . . . Peinado, R. (2001). An analysis of phonological awareness instruction in four kindergarten basal reading programs. *Reading and Writing Quarterly, 17,* 25–51.

Smith, S. B., Simmons, D. C., & Kame'enui, E. J. (1998). Phonological awareness: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics* (pp. 61–128). Mahwah, NJ: Lawrence Erlbaum.

Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at risk for reading failure. *Journal of Learning Disabilities, 38,* 387.399.

Stahl, S. A., Duffy-Hester, A. M., & Dougherty Stahl, K. A. (1998). Everything you wanted to know about phonics (but were afraid to ask). *Reading Research Quarterly, 33*, 338–355.

Stahl, S. A., & Murray, B. (1998). Issues involved in defining phonological awareness and its relation to early reading. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 65–87). Mahwah, NJ: Lawrence Erlbaum.

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

Sweet, R. W. (2004). The big picture: Where we are nationally on the reading front and how we got here. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 13–44). Baltimore, MD: Paul H. Brookes.

Teddlie, C., Reynolds, J., & Sammons, P. (2000). The methodology and scientific principles of school effectiveness research. In C. Teddlie & D. Reynolds (Eds.) *The international handbook of school effectiveness research (pp. 55-133*). New York, NY: Falmer.

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

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe

reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33–58, 78.

Torgesen, J. K. (1998). Catch them before they fail: Identification and assessment to prevent reading failure in young children. *American Educator, 22,* 32-39.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 148–173). Mahwah, NJ: Lawrence Erlbaum.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1*, 217–234.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvin, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*(4), 579–593.

Treiman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development, 69*, 1524–1540.

University of Oregon Center on Teaching and Learning and Oregon Department of Education (2011). Professional development portal for the Oregon K-12 literacy framework. Retrieved from: http://oregonliteracypd.uoregon.edu/.

U.S. Department of Education. (2008a). *Reading First funding status.* Retrieved from http://www2.ed.gov/programs/readingfirst/funding.html

U.S. Department of Education. (2008b). *Reading First: Student achievement, teacher empowerment, national success.* Retrieved August 15, 2010 from http://www.ed.gov/nclb/methods/reading/readingfirst.pdf.

U.S. Department of Education & National Institute of Child Health and Human Development. (2003). *National symposium on learning disabilities in English language learners: Symposium summary.* Washington, DC: Authors.

Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of Nonsense Word Fluency for English language learners. *School Psychology Review, 37*, 1–13.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Pedrotty Bryant, D., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*, 301–315.

Vellutino, F. R. (1991). Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. *Journal of Educational Psychology, 83*, 437–443.

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention with low response to intervention. *Journal of Learning Disabilities, 41*, 126–142.