

EVALUATING THE RELIABILITY OF SELECTED SCHOOL-BASED INDICES OF
ADEQUATE READING PROGRESS

by

COURTNEY E. WHEELER

A DISSERTATION

Presented to the Department of Special Education
and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2010

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Courtney Wheeler

Title:

"Evaluating the Reliability of Selected School-Based Indices of Adequate Reading Progress"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Roland Good, Chairperson, Special Education and Clinical Sciences

Edward Kameenui, Member, Educational Methodology, Policy, and Leadership

Kenneth Merrell, Member, Special Education and Clinical Sciences

Sara Hodges, Outside Member, Psychology

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

June 14, 2010

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

© 2010 Courtney E. Wheeler

An Abstract of the Dissertation of
Courtney E. Wheeler for the degree of Doctor of Philosophy
in the Department of Special Education and Clinical Sciences
to be taken June 2010

Title: EVALUATING THE RELIABILITY OF SELECTED SCHOOL-BASED
INDICES OF ADEQUATE READING PROGRESS

Approved: _____
Dr. Roland H. Good, III

The present study examined the stability (i.e., 4-month and 12-month test-retest reliability) of six selected school-based indices of adequate reading progress. The total sampling frame included between 3970 and 5655 schools depending on the index and research question. Each school had at least 40 second-grade students that had complete Oral Reading Fluency (ORF) data for the time periods in question. Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores were used to examine school-wide adequate reading progress. The stability of those indices from semester-to-semester and from year-to-year across the 2005-2006, 2006-2007, and 2007-2008 school years was examined. Adequate progress was defined as students improving their instructional recommendation

(i.e., reducing their level of risk) or remaining at benchmark (i.e., remaining on track) over a specified period of time. The six indices were as follows: (1) outcome percent established, (2) percent adequate progress, (3) intensive percent adequate progress, (4) strategic percent adequate progress, (5) barely benchmark percent adequate progress, and (6) school-wide high rates of adequate progress. The indices were intended to provide a snapshot of how well a school's reading instruction is meeting student needs.

Based on the analysis, the stability coefficients ranged from .10 to .90 indicating that certain indices had higher stability coefficients than others. Overall, the year-to-year indices tended to be more stable than the semester-to-semester indices. Between 143 and 203 schools had school-wide high rates of adequate progress over a two- and three-year period. These findings indicate that schools can be generally effective in helping their students achieve high rates of adequate reading progress and that schools are able to maintain and support high rates of adequate progress for consecutive cohorts of second-grade students. Results are discussed within a broader framework of school effectiveness indices and response to intervention.

CURRICULUM VITAE

NAME OF AUTHOR: Courtney E. Wheeler

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon; Eugene, OR
Rochester Institute of Technology; Rochester, NY
St. Lawrence University; Canton, NY

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, 2010, University of Oregon
Master of Science and Advanced Graduate Certificate, School Psychology, 2007,
Rochester Institute of Technology
Bachelor of Arts, Psychology and Economics, 2004, St. Lawrence University

AREAS OF SPECIAL INTEREST:

Response to Intervention
Early Literacy
Systems-level Changes in Schools
School-wide Models of Support

PROFESSIONAL EXPERIENCE:

School Psychologist
Heartland Area Education Agency, Johnston, IA, Aug 2009 - present
Lane Education Service District, Eugene, OR, Jan 2008 - March 2008
Syracuse City School District, Syracuse, NY, Aug 2006-June 2007

Graduate Teaching Fellowship, University of Oregon, 2007-2009

Research Team Member, DIBELS/IDEL Research Team, University of Oregon,
Eugene, 2007-2009

Protocol Reviewer, Rochester Institute of Technology, Rochester 2005-2006

Psychology Teaching Assistant, St. Lawrence University, Aug 2002- Dec 2002

GRANTS, AWARDS AND HONORS:

DIBELS Student Support Award, University of Oregon, 2007-2009

Oregon State Association of School Psychologists (OSPA) Student
Representative, 2008-2009

School Psychology Program General Travel Funds Award, University of Oregon,
2007

Graduate Class Representative, Rochester Institute of Technology, 2004-2007

ACKNOWLEDGMENTS

I wish to express my sincere gratitude and appreciation first to Professor Good for his assistance and support throughout this process. He provided sound advice, timely feedback, and professional judgment in helping to formulate this dissertation from the beginning to end. His support and enthusiasm helped to see this project through and I am thankful for his commitment. I also must thank Dr. Kame'enui and the Center on Teaching and Learning in the College of Education at the University of Oregon for assistance with obtaining the limited access dataset from the DIBELS Data System for this study, as it would not have been possible without that support.

The DIBELS Data System (DDS) is operated and maintained by the Center on Teaching and Learning (CTL) in the College of Education at the University of Oregon in collaboration with state, regional, and local educational agencies. This manuscript was prepared using a limited access dataset obtained from the DDS. The author is solely responsible for the content of this manuscript, which does not necessarily reflect the opinions or views of the DDS, the University of Oregon, or DDS users.

For my husband, Pavel, without whose caring support this work would not be possible, and my parents, Keith and Karen, who passed on a love of reading and respect for education.

TABLE OF CONTENTS

Chapter	Page
I. PROBLEM STATEMENT	1
II. LITERATURE REVIEW	7
School-Wide Systems for Academic Support	7
Response to Intervention	10
Role of Formative Assessment within a School-wide System	14
Indices of School Academic Effectiveness	18
III. METHODS	21
DIBELS Data System	21
Participants	22
Measures	23
Dynamic Indicators of Basic Early Literacy Skills (DIBELS).....	23
Design and Procedure	23
Outcome Percent Established	25
Percent Adequate Progress	26
Intensive Percent Adequate Progress	28
Strategic Percent Adequate Progress	29
Barely Benchmark Percent Adequate Progress	30
School-wide High Rates of Adequate Progress	31
Next Steps	33
Data Analysis	34
Summary	35
IV. RESULTS	36
Descriptive Statistics	37
Research Question 1	40

Chapter	Page
Research Question 2	56
Research Question 3	61
Summary of Results	64
V. DISCUSSION	66
Stability of Indices	67
Practical Implications for Three Tiered Models	71
Limitations	72
Future Research	75
Conclusion	78
REFERENCES	80

LIST OF FIGURES

Figure	Page
1. Three Tiered System of Support	8

LIST OF TABLES

Table	Page
1. Descriptive Statistics for Second Grade by Outcome between Cohort Year and Index Time Period	38
2. Descriptive Statistics for Second Grade by Initial Status Outcome between Cohort Year and Index Time Period	40
3. Outcome Percent Established: Stability Coefficients of 2 nd Grade First Semester to Second Semester	42
4. Second Grade, Semester-to-semester Stability Coefficients for Percent Adequate Progress	44
5. Second Grade, Semester-to-semester Stability Coefficients for Intensive Percent that Make Adequate Progress.....	46
6. Second Grade, Semester-to-semester Stability Coefficients for Strategic Percent that Make Adequate Progress	49
7. Second Grade, Semester-to-semester Stability Coefficients for Barely Benchmark Percent that Make Adequate Progress	51
8. Second Grade, Semester-to-semester Stability of Decisions with Schools that have System Wide High Rates of Adequate Progress.....	55
9. Second Grade, Year-to-year Stability Coefficients by Index	57
10. Second Grade, Year-to-year Stability of Decisions with Schools that have System Wide High Rates of Adequate Progress	60
11. Second Grade, Semester-to-year Correlation Coefficients by Index.....	63
12. Second Grade, Year-to-semester Stability of Decisions with Schools that have System Wide High Rates of Adequate Progress	64

CHAPTER I

PROBLEM STATEMENT

Recent results from the National Assessment of Educational Progress (NAEP) indicated that the average reading scores for fourth grade students has risen 4 points in seventeen years and there have been no changes in the national scores from 2007 to 2009 ((National Center for Education Statistics, 2009). According to the NAEP results, 67% of the students in the US sample are performing at or above the basic level. Alarming, this is only a 5% increase in seventeen years. This indicates that 33% are not meeting the most basic level of reading proficiency on the NAEP. This is after a large amount of public attention has been given to reading and reading development in the past decades. The NAEP's definition of basic level of reading proficiency consists of fourth grade students being able to "locate relevant information, make simple inferences, and use their understanding of the text to identify details that support a given interpretation or conclusion. Students should be able to interpret the meaning of a word as it is used in the text" ((National Center for Education Statistics, 2009, p. 18). Countries, districts, and schools may be doing an injustice to their students by saying that passing at the basic level is "good enough".

When looking at the sample of students that are performing at or above the proficient level on the NAEP, the outlook for reading scores in the United States grows increasingly dim. NAEP defines reading at a proficient level as a student being able to

“integrate and interpret texts and apply their understanding of the text to draw conclusions and make evaluations” (National Center for Education Statistics, 2009, p. 18). The 2009 NAEP results report that only 33% percent passed at or above the proficient level. This was only 4% increase in seventeen years. This indicates that 67% of students are not meeting grade level standards for fourth grade. The outlook is even more dismal when looking at minority populations and those of low socio-economic status; 84% of African-American students, 83% of Hispanic students, and 83% of those eligible for the national school lunch program were below the proficient level in 2009.

It is imperative that the education system in the United States make improvements to meet the educational needs of all students. Schools need to ensure that they are meeting the needs of their students through differentiated instruction at differentiated levels of support (i.e., all three tiers of support-universal, secondary, and tertiary) and ensure that the efforts that they are putting forward are generally effective for all their students. Schools need to have a reliable and efficient snapshot of school-based adequate reading progress, ensuring an ability to evaluate if the school is providing and maintaining quality education for all.

Providing that efficient snapshot of school-based adequate reading progress requires an understanding of the structures that exist in schools and the constructs behind those structures. These constructs are often intertwined and share similar features. The three constructs that this study draws from were: (a) three tiered systems of support, a model which uses differentiated levels of instruction and differentiated levels of support including a universal level, a secondary level, and a tertiary level of support; (b)

Response to Intervention (RTI), which links student progress to different tiers of instructional support and can be used for special education eligibility; and (c) school effectiveness indicators, which may include end of the year outcomes, annual yearly progress that is either dynamic or static, and the proposed indices of adequate reading progress.

Three tiered systems of support and RTI use differentiated tiers of instruction and levels of support in order to support all students within a school. The maximization of learning opportunities for all students is a key component of these constructs. The different tiers of instruction in RTI and three-tiered systems of support allow students to receive differentiated support across tiers in order to receive the instruction that is best suited to their educational needs. With reading, these systems provide universal or “core” support for all students in the first tier and then build intervention curriculums with increasing levels of support for students that require additional reading instruction to be successful. However, how do schools measure if their tiers of support are effective in teaching reading to their students?

Little research has been conducted specifically on school academic effectiveness indices. However, the need to identify possible indices of school academic effectiveness exists. These indices are needed to evaluate the different tiers of support within RTI. Indices could also provide an efficient snapshot of different aspects of the school, such as reading progress within a grade or building. Some possible indices that have been considered include end-of-the-year outcomes, annual yearly achievement progress, and additional indices that are not widely discussed in the literature.

One possible index of school-wide academic effectiveness is the end of the year outcomes that are measured on a state-level assessment test. The higher number of students a school has that passed the assessment, then the higher the school is rated. A drawback to an end of the year outcomes index is that initial skills strongly predict outcomes. Therefore, a school that begins the school year with many highly skilled students would be more likely to achieve higher outcomes at the end of the year. To address these concerns, a progress index can be used. A second possible index is annual yearly progress as measured by No Child Left Behind, which follows a similar line of logic to the end of the year outcomes, but takes different factors into account based on the state requirements.

Overall, there are few indices that assist the school in determining if the instruction that they are providing as part of a RTI framework or three-tiered system of support is effective in meeting the needs of all students. This lack of research is system wide and encompasses behavior and academics. This dissertation focused on one dimension of the academic component, specifically on different school-based indices of adequate reading progress. The primary goal of this dissertation was to investigate six proposed indices of school-based adequate reading progress and examine the reliability, or stability, of these indices within and across three academic years for schools for the same grade level. The particular indices, as discussed in detail in the methods section, were chosen because it was hypothesized they may indicate if a school is meeting the reading needs of students with different instructional recommendations. Each school had at least 40-second grade

students, the equivalent of two classrooms in a school, which had complete Oral Reading Fluency (ORF) data for the time periods in question.

The six indices were created to examine the stability of Dynamic Indicators of Basic Early Literacy Skill (DIBELS) scores of second grade students from semester-to-semester and from year-to-year across the 2005-2006, 2006-2007, and 2007-2008 school years. The indices were as follows: (1) outcome percent established, (2) percent adequate progress, (3) percent intensive adequate progress, (4) percent strategic adequate progress, (5) percent barely benchmark adequate progress, and (6) school-wide high rates of adequate progress. Definitions of the indices occur in the method section.

Two of the indices, outcome percent established and percent adequate progress, were representative of all students in the schools in the sample (those initially scoring within intensive, strategic, barely-benchmark, and well-above benchmark DIBELS recommendation levels). Intensive percent adequate progress, strategic percent adequate progress, and barely benchmark adequate progress indices were representative of a specific subset of the student population within schools in the sample. In contrast, system-wide high rates of adequate progress was representative of the students that initially scored in the intensive, strategic, and barely benchmark ranges and were in schools that had high rates of helping their students achieve adequate progress. The specific research questions were:

1. Which indices provided the most stable information about school based adequate reading progress from first semester to second semester?

2. Which indices provided the most stable information from the first year, 2005-2006, to the second year, 2006-2007, to the third year, 2007-2008?
3. How do the semester indices relate to the yearly indices?

CHAPTER II

LITERATURE REVIEW

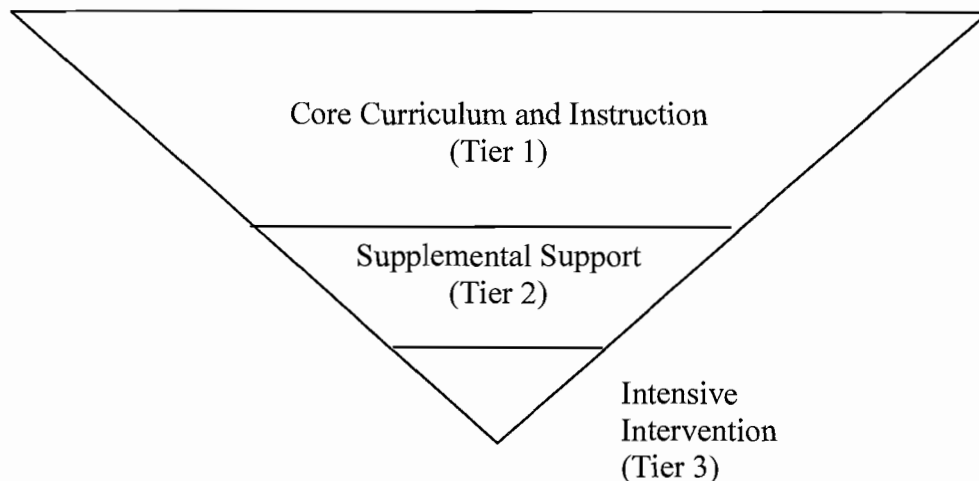
Research on effective school-based systems has been addressed for many decades in the educational field. The necessity of having academic systems within schools that meet student needs is a given. However, a limited amount of research has been conducted to evaluate the effectiveness of those academic systems through different school based indices of academic effectiveness, specifically on indices of adequate reading progress. This literature review section has three goals, the first of which is to define school-wide systems for academic support and discuss the delivery of support within schools. Specifically, a three-tiered model of support is discussed in the context of RTI. The second goal of the literature review is to summarize the role of formative assessment within a school-wide system. Finally, the research base for possible indices of school academic effectiveness is examined. This review is not intended to be exhaustive, but instead provides a basic framework for understanding the literature base related to the current study.

School-Wide Systems for Academic Support

School-wide systems allow for differentiated levels of support for academic instruction. The school-wide system that is frequently written about in scholarly research is a multi-tiered system of support. Typically, this system gives guidance to school on a structure to provide differentiated academic instruction at each of the different tiers of

support. McIntosh, Chard, Boland, and Horner (2006) described the principles of the multi-tiered model succinctly. This model includes: (a) universal intervention for all students, (b) a system of screening to determine appropriate services for students, and (c) implementation of a continuum of services based on the students' needs. Kratochwill, Volpiansky, Clements, and Ball (2007) stated that multi-tiered systems typically encompassed three tiers: (a) primary intervention, (b) supplementary intervention, and (c) intensive intervention supports. The three-tiered school wide-system was adopted from the three-tiered public health services model to describe levels of support in the academic and social domains (Merrell & Buchanan, 2006). Maximizing student outcomes is a main consideration within the three-tiered system of support. Visually the three-tiered support system can be seen in Figure 1.

Figure 1. Three Tiered System of Support



In the first tier, the universal, or core, curriculum should be generally effective in teaching the majority of students within a school (Fuchs & Fuchs, 2006). The core curriculum should be preventative of student failure, proactive with addressing student needs, and involve all students in the school. With academics, this can be conceptualized as teaching a core curriculum that has a strong research base with high levels of efficacy (Daly, Martens, Barnett, Witt, & Olson, 2007). Within this tier, formative assessments are used as a system of screening to determine appropriate support services for students (Gresham, 2004; Daly et. al. 2007). The students that are at risk for academic problems should then receive supplementary services in the second tier of support (Fuchs & Fuchs, 2006). This supplemental support consists of additional instruction and continued progress monitoring in order to ensure that students are progressing toward goals. This instruction is usually conducted in a small group setting. If sufficient progress is not being made, then the student should receive intensive intervention, or tier three supports (Fuchs & Fuchs, 2006). These supports are individualized, tailored instruction and continued progress monitoring should occur. As each tier increases, the instruction becomes more specifically tailored to individual student needs and the grouping within that instruction decreases (Vaughn & Fuchs, 2003; Reschly, 2005). The three-tiered system of support is an essential component in RTI as is the movement between the tiers of intervention based on the ongoing progress monitoring data (Kratochwill et al., 2007).

Evaluating the quality of the three-tiered system of support is vital. Schools need to be able to determine if each tier is meeting the needs of the students that are receiving services within it in order to determine if changes in instruction are necessary. If the

curriculum in the first tier is not generally effective for the majority of students, then the health of the overall system needs to be considered. The supplemental support at the second tier and intensive support at the third tier also need to be generally effective for the students within those tiers. Prior to being able to make educationally significant decisions about students, the overall academic effectiveness of the school needs to be addressed. Only after this occurs can decisions, such as which level of support for a particular student is needed or if a specific learning disability eligibility decision is warranted with RTI, be made.

Response to Intervention

Daly et al. (2007) reported that the first conceptualization of RTI could be traced back to a report on the overrepresentation of minorities in special education from 1982 by the National Academy of Sciences (Heller, Holtzman, & Messick, 1982). Historically, according to Daly et al (2007), the origins of RTI can be traced back even further into the traditions of applied behavior analysis from the 1960's. For the purposes of this dissertation, the definition of RTI by the National Association of State Directors of Special Education (NASDSE) is used. NASDSE defines RTI as “the practice of providing high-quality instruction and interventions matched to student need, monitoring progress frequently to make decisions about changes in instruction or goals and applying child response data to important educational decisions” (Batsche, Elliott, Graden, Grimes, Kovalski, Prasse, Reschly, Schrag, & Tilly, 2005, p.3). Thus in evaluating response to intervention models it is critical to have a method for evaluating the quality of instruction. This could be accomplished either on an individual basis or by examining an

index of school-wide adequate reading progress. This dissertation examines indices that may be used to evaluate school-wide adequate reading progress so a school can determine if generally effective reading instruction is in place.

Fuchs and Fuchs (2006), among other researchers, conceptualized the three-tiered system as a part of RTI. The theoretical framework has been widely discussed in the field of education. RTI can be conceptualized at least in two main ways. It can be thought of as one way to provide interventions to students who are struggling or as method for determining special education eligibility for learning disabilities. Maximizing learning opportunities for all students is a key component of RTI. The different tiers of instruction in RTI allow students to move across tiers to receive the instruction that is best suited to their educational needs. If implemented correctly, RTI is an effective educational framework for all the students in a school.

Two approaches to providing interventions with RTI have been discussed in the educational literature (Fuchs, Fuchs, & Compton, 2004). The standard protocol approach involves utilizing research-based interventions, which provide effective instruction for the majority of students participating in the interventions. Standard-protocol interventions are typically small group interventions involving a smaller number of students than the core curriculum instruction. Daly et al. (2007) described this approach as a common strategy for delivering interventions; however they cautioned that it may not work for all students. An alternative approach involves using a problem-solving framework. The intensity of instructional need that a student exhibits is matched with the intensity of the intervention. A hybrid of these approaches could also be used in schools, especially those

with limited resources in order to effectively address students' needs at different tiers of instruction. A standard protocol approach could be used for students needing second tier support and problem-solving approach could be utilized with the students that need specialized support in the third tier. The problem solving approach would be used after it is determined that the standard protocol approach was implemented with fidelity and the student required increasingly intensive interventions to make adequate progress.

Both the problem solving and standard protocol approaches fit into a three-tiered model of academic support. Fuchs and Fuchs (2006) conceptualized the first tier, primary prevention, as the instruction that occurs in the general education classroom. The students that do not respond to the first tier research-based core instructional program would then be given additional support at the secondary tier. Fuchs and Fuchs (2006) described the secondary tier as scientifically validated small group instruction. They conceptualized the tertiary tier as an individualized program that is tailored to the needs of the student in question. Fuchs and Fuchs (2006) noted that special education should be a resource within the prevention system. Students would enter and exit special education as needed. However, Fuchs and Fuchs (2006) do not conceptualize special education as a specific tier in the three-tiered model.

As many school districts across the country strive to adopt RTI practices, questions have arisen about how to intervene with students that are struggling academically and, if necessary, how to determine special education eligibility. In 2004, the re-authorization of the Individuals with Disabilities Education Improvement Act (IDEA) no longer required a severe discrepancy between achievement and ability to be

the only option in determining special education eligibility for Specific Learning Disabilities (IDEA, 2004). Instead, the law states that educators are also allowed to use a process that determines if the child responds to scientifically, research-based intervention as part of the evaluation procedures (IDEA, 2004). As a result, attention from the field of education has focused on RTI as a means for determining special education eligibility in the area of Specific Learning Disabilities.

RTI exists as a theoretical framework in which researchers have discussed a continuum of services for all students. Gresham, VanDerHeyden, and Witt (2005) conceptualized learning disabilities within the RTI model as: (a) representing a low level of performance, (b) slow growth rate despite receiving evidence based instruction, (c) an adverse impact on the education of the student, a demonstrated need for special education services, and (d) specific goals for exiting the student from special education. Having evidence-based instruction and intervention is an essential feature of being able to identify learning disabilities within the RTI model. Fuchs and Fuchs (2006) suggested that for each tier (universal, supplementary, and intensive), questions be asked regarding the intervention efficacy, assessment integrity, and feasibility of the intervention.

Ultimately, if the three-tiered model of support contains effective and relevant interventions that are feasible within a school and delivered with integrity, student achievement will be higher (Mellard & Johnson, 2008). According to Mellard and Johnston (2008), this is because of the proactive process with data collection, analysis, and data-based decision making associated with RTI. Daly et al. (2007) cautioned that without high quality intervention programs and sequential decision-making processes,

breakdowns in service delivery could occur causing a delay in the delivery of effective interventions to students. Fuchs and Fuchs (2006) stipulated that teachers could use the data collected to aid with instructional decisions for specific students and entire classes. School-wide data, through formative assessment techniques, can be used to determine the health of the overall system.

Role of Formative Assessment within a School-wide System

Formative assessment is a process in which “curriculum based measurement” (CBM) is used to determine academic intervention effectiveness (Deno, 1986). Formative assessment began as a way to measure an individual student’s progress and has evolved into a way of providing information on school-wide systems. This transition has allowed educators to investigate the different tiers of their school-wide structure and modify their instructional practices if needed.

CBM is one of the most widely studied forms of formative classroom assessment (Fuchs, Fuchs, & Compton, 2004). This type of assessment allows educators to assess students with an instrument that is reliable, valid, and efficient to administer (Deno, 1985). At a school level, the CBM data can be used to evaluate if the instructional environment is meeting students’ needs within a three-tiered system. Each tier can be examined to ensure that the majority of students within that tier are making adequate progress toward instructional goals that have been set. If the majority of the students in each tier are not making adequate progress, then modifications in the instruction must occur.

At the student level, CBM also allows educators to both view the student at a certain point in time and the progress that the student makes across a period of time. It can also be used to compare a particular student to a classroom and a classroom to a grade level performance criterion. CBM can be distinguished from other classroom assessments in many ways. It is standardized with documented reliability and validity. CBM can be conducted frequently and over a long period of time. It also has alternate forms, which is essential for a progress-monitoring device. CBM serves to create optimal learning conditions for teachers and students by providing corrective feedback for both (Gravois & Gickling, 2002). CBM is used to assess skills as a student enters the classroom and provides insight to interventions around which instruction can be designed at an individual student level. It is also used to evaluate the effectiveness of the instruction over time.

One of the most widely researched and implemented CBM tasks are that of oral reading rate (R-CBM). Adams (1990) referred to the speed in which text is spoken out loud as the most important characteristic of reading. This is oral reading fluency, which is the oral reading of text with speed and accuracy (Fuchs, Fuchs, Hosp, & Jenkins 2001). R-CBM involves counting the number of words read correctly from meaningful connected text in a 1 minute time period. Because it looks at words read correctly in a given time frame, it is sensitive enough to detect small changes in reading proficiency (Fuchs & Fuchs, 1986). R-CBM can be used to generate time-series sets of data to be displayed graphically and can result in better instruction through higher expectations and more adaptations to instructional programs (Fuchs, Fuchs, & Hamlett, 1989). The R-

CBM also is related to the reader's skill at making meaningful connections between sentences, relating text meaning to a reader's knowledge, and making inferences to complete the story (Fuchs et. al, 2004).

A specific type of CBM for reading is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) developed by Kaminski and Good (1996). DIBELS are a series of measures designed to assess a student's fluency in fundamental reading skills. The DIBELS Oral Reading Fluency (DORF), a subtest of DIBELS, assesses the number of words that the student can read correctly in a 1 minute time period. In that one minute, the student is asked to read as quickly and correctly as he or she can. The examiner records the number of words omitted, substituted, and hesitations for more than three seconds as errors and at the end of one minute records the number of words read. The oral reading fluency rate is then calculated as the number of words read correctly in one minute (Kaminski & Good, 1996). Benchmark assessments using DORF, typically conducted three times per year, are a component of formative assessment. After conducting benchmarking assessments, students in need of extra academic support would be identified based on their assessment results.

Martson (1989) reviewed the early reliability and validity of ORF in Curriculum-Based Measurement. He found that the overall test-retest reliability ranged from .82 to .97. Martson also found that alternate forms and inter-rater reliability had correlation coefficients above .90. This indicates that ORF is an acceptable assessment tool. The passages on the DIBELS - ORF were designed to be consistent with the Test of Reading Fluency (TORF) (Children's Educational Services, 1987) in readability, reliability and

validity (Kaminski & Good, 1996; Good, Wallin, Simmons, Kame'enui, Kaminski, 2002). The TORF consists of standardized passages and administration procedures that can be used as a screening device for children and measure growth in reading skills (Children's Educational Services, 1987). Concurrent validity with the TORF passages ranged from .92 to .96 (Good et. al, 2002).

Numerous studies have been conducted comparing DORF and the relationship to various state high stakes tests that satisfy NCLB legislation. For example, Wilson (2005) found that with third-grade students who achieved the benchmark on the DORF, the measure was likely to identify those students with sufficient accuracy for passing the Arizona Instrument to Measure Standards. For the students who were considered to be at risk when given the DORF, the measure could identify those who were unlikely to meet the proficiency standard on the Arizona Instrument to Measure Standards. The DORF measure could be used equally well for various demographic subgroups that differed in gender, ethnicity, and income level. Shaw and Shaw (2002) used DORF to examine the relationship between those measures and performance on the Colorado State Assessment Program (CSAP) involving third-grade students. The third-graders who read 90 or more words correctly in one minute on the DORF were 91% likely to score "proficient" or "advanced" on the CSAP, and those that read less than 90 on the DORF were more likely to receive a score of "unsatisfactory" or "partially proficient" (Shaw & Shaw, 2002). Buck and Torgesen (2003) investigated the relationship between the CBM and the Florida Comprehensive Assessment Test (FCAT) and found that there was a significant correlation between CBM scores and reading FCAT scores. Another study found DORF

an accurate predictor of whether a student would achieve a proficient score on the North Carolina End of Grade Reading Assessment (Barger, 2003).

Overall, formative assessment, including DORF, is a valid and reliable measure that can provide information regarding the students' progress toward literacy goals within a school. It can also be used to provide a snapshot of school-wide academic, specifically reading effectiveness over the different tiers of the school-wide system. However, the question arises about how information from formative assessments, such as DORF, can be turned into stable indices that focus on providing that snapshot for academic (i.e., reading) effectiveness. A number of possible indices exist, but they must be evaluated to determine if they provide the necessary information to be useful in informing schools about their reading progress at all tiers.

Indices of School Academic Effectiveness

Schools must determine if they have effective instructional strategies, programs, approaches, or techniques to support their students. The key features of any type of evaluation tool that would be used to evaluate instructional programs are that they must be reliable, valid, and meaningful at the school level. However, the research base in school-based indices of academic achievement is lacking. There are few constructs in the research that could be considered possible indices and each potential index has numerous advantages and disadvantages.

With the emergence of No Child Left Behind (2001), schools are required to utilize high stakes tests to assess student achievement. NCLB was created to strengthen accountability in the United States school system by requiring states to establish

challenging standards in reading and mathematics, conduct annual testing for all students in grades 3-8, and meet annual standards (NCLB, 2001). As a result, most of the research on indices of academic effectiveness currently focuses on end of the year, high stakes test results as a measure of school effectiveness.

Linn (2008) discussed the methodological differences between status models, such as yearly progress targets from NCLB, and growth models, such as state accountability standards. Carlson (2002), as cited in Linn (2008, p. 700), described status as “how good is a school” and change or growth as “is a school getting better”. Another method of conceptualizing this would be the status model looks at the school at one point in time, a static measure, and the growth model looks at the school over a period of time, a dynamic measure. A status approach to school accountability would measure the current levels of achievement against the performance targets that were set by NCLB and would address school effectiveness based on the achievement level of its students. In contrast, a change or growth approach would determine if student achievement is improving and evaluate school effectiveness by determining if the effectiveness of the overall school is improving. Linn (2008) states that for the change model, the measurement of growth requires longitudinal tracking of achievement scores from cohorts of students instead of grades. The growth models take into consideration differences that exist in each school before the test is administered.

Linn (2008) also identified a subset of the growth models as the value-added model. Within a value-added model, the contribution that schools or teachers make to student achievement would be identified. Linn described this as investigating student test

data across a period of a few years. However, the research that he cited cautions against the use of the value-added model to determine the effects of instruction. Overall, Linn suggested that accountability systems should be used to identify schools that need more intensive interventions to help their students rather than as a system to sanction schools. Duran (2005) suggested that the NCLB legislation created an opportunity for states to identify effective schools and be able to provide support to the schools that need improvement.

Linn, Baker, and Betebenner (2002) stated that the adequate yearly progress targets of NCLB do not take into account the differences from year to year in different cohorts of students. Thus, it is unclear how a school would determine if the academic instruction that students receive were addressing the needs of the students within the school as a whole. One approach to evaluate the effectiveness of instruction is to observe and rate the intervention. A second approach is to evaluate the system of instruction in the school using formative assessment measures that are already given three times per year. Different indices of academic progress might be considered for this purpose. Indices could be varied in the type of snapshot they provided the school. The present study examined six indices that are based on formative assessment measures so that schools could potentially have access to a stable and efficient snapshot of the school-based adequate reading progress.

CHAPTER III

METHODS

This study utilized archival data from an intact data set from the DIBELS Data System, as described in the Data Source section. The participants' section outlines the requirements for a school to be included from the 2005-2006, 2006-2007, and 2007-2008 academic years. DIBELS Oral Reading Fluency data were utilized for this study and are described in the measures section. The design and procedure section outlines the safeguards for protecting the students' and schools' identity in the archival data sample. This section also outlines the data usage agreement through the Center on Teaching and Learning and the University of Oregon's Internal Review Board. Detailed steps for acquiring the data and computing each of the six indices are discussed, followed by a description of the programs that were written. In the data analysis section, the correlational and Kappa analyses required for each the three research questions are described.

DIBELS Data System

The DIBELS Data System (DDS) is a web-based database that is housed at the University of Oregon and operated by the Center on Teaching and Learning (CTL) since 2001. According to the DDS description, over 15,000 schools utilize the data system to enter student performance results and create reports based solely on student DIBELS

scores. The data system reports can be used to track and measure progress of students, classes, schools, and districts (DDS, n.d.).

The DDS is organized by school district. Each district has a coordinator that is responsible for assigning schools and school-level coordinators to evaluate the data. At the school level, the DDS is organized to best suit each schools' needs, providing information to facilitate decisions at the school, classroom, intervention group, and individual student levels. According to DDS (n.d.), student records are typically organized by classrooms, but can also be assigned to a specific instructional group. The DDS allows users to input benchmark DIBELS data, typically three times per year, and progress monitoring assessment data, typically weekly, bi-weekly, or monthly, for grades kindergarten through sixth (DDS, n.d.). This includes the measure that is specific to this study, DIBELS Oral Reading Fluency, and the other DIBELS measures when indicated (i.e., Nonsense Word Fluency, Retell Fluency, Word Use Fluency, Initial Sound Fluency, Letter Naming Fluency, and Phoneme Segmentation Fluency). Users can also input demographic information about students and scores from an end of the year state outcome test. The DDS has features that allow users to generate reports based on the data at the student, class, school, district, and project level (DDS, n.d.).

Participants

Participants were all from schools that had 40 or more second grade students, equivalent to about two classrooms, with complete data for DIBELS Oral Reading Fluency at the beginning, middle, and end of the year in the DDS for the 2005-2006, 2006-2007, and 2007-2008 school years. Second grade was chosen as the sampling year

because it is the first year that DIBELS Oral Reading Fluency is given during all three benchmarking periods (i.e., fall, winter, and spring). The total sampling frame included between 3970 and 5655 schools nationwide depending on the research question and index being analyzed. The sampling size varied due to the number of schools with 40 or more students that had complete data, specifically data for the time periods in question (e.g. first semester to second semester). The participating schools included only those that utilize the DDS to enter benchmarking data specific to DIBELS.

Measures

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

A component of DIBELS is DIBELS Oral Reading Fluency (DORF). DORF is a curriculum based measure of students' ability to accurately and fluently read connected text. The DORF passages are standardized and have "administration procedures designed to (a) identify children who may need additional instructional support, and (b) monitor progress toward instructional goals" (Good & Kaminski, 2002, p. 30). A student is asked to read from a grade level passage for one minute. The number of words read correctly in that one minute time period is the oral reading fluency rate. Validity estimates range from .67 to .96 and reliability estimates for DORF range from .89 to .96 (Good, Simmons, & Kame'enui, 2001; Good & Kaminski, 2002).

Design and Procedure

This study was conducted using archival data from the DIBELS data system. The DORF data were de-identified and the DDS provided a unique identification number to each student and school. The researchers did not have access to any identifying data or

information, including school names, demographics, identifiable student information, etc. The data set was obtained after completing the Center on Teaching and Learning's data usage application and the University of Oregon's Internal Review Board process. The researcher complied with the recommendations of both processes. The data set included the fall, winter, and spring benchmark scores for DORF. Students with incomplete data on DORF were removed from the data set. For example, if the analysis in question included first semester data, then a student who did not have beginning of year or middle of year data on DORF was removed from the data set. Schools with less than 40-second grade students with complete data on DORF were excluded from analysis.

The data was read into SAS and programs were created to identify the schools with 40 or more students and assign each student to a risk status level: high risk (i.e., intensive), some risk (i.e., strategic), and low risk (i.e., benchmark), for each benchmark period. The low risk category was divided into barely benchmark, which was within 20 words read correct above the benchmark goal, and a well above benchmark category, which was 20 words or more read correctly above the benchmark goal. Within 20 points of the benchmark goal was chosen because it is likely that the students scoring in this range are at greater risk for not remaining within the benchmark category and therefore have an increased risk of not meeting end of the year standards.

The recommendation levels utilized for the second grade sample were obtained from the DIBELS benchmark goals and indicators of risk and are as follows: (a) intensive students read less than 26 words correct per minute at the beginning of the year, less than 52 at the middle of the year, and less than 70 at the end of the year; (b) strategic students

read between 26 and 43 words correct per minute at the beginning of the year, between 52 and 67 at the middle of the year, and between 70 and 89 at the end of the year; and (c) benchmark students read 44 or more words correct per minute at the beginning of the year, 68 or more at the middle of the year, and 90 or more at the end of the year. These recommendation levels increase in difficulty from the beginning of the year, to the middle of the year, and to the end of the year. Therefore students have to be making growth in order to remain within the same instructional recommendation level that they started the semester.

Based on the results from the first set of analyses, SAS programs were written to analyze the following possible indices of adequate reading progress as outlined in the significance section. The indices further described below were as follows: (1) outcome percent established, (2) percent adequate progress, (3) intensive percent adequate progress, (4) strategic percent adequate progress, (5) barely benchmark percent adequate progress, (6) school-wide high rates of adequate progress.

Outcome Percent Established

The first index was the percentage of students who reached the benchmark goal at the middle of the year or end of the year, depending on the research question. A SAS program was created that examined all of the 4039-5655 schools for whom this index was computed. First, the program obtained the count of students in the school at all instructional recommendation levels at the beginning of the sampling period that had complete data for the assessment time period. The program then created a count of students that reached or maintained their benchmark status. The percentage of students

that began at any instructional recommendation point and reached or remained at benchmark was assigned to each school. This was calculated by taking the number of students who scored in the benchmark range at the end of the assessment period and then divided by the total number of students at the beginning of the assessment period. This was completed for each school. For example, if a school with complete data at the beginning of the assessment period had 13 students in the intensive range, 40 students in the strategic range, 15 students in the barely benchmark range, and 20 students in the well-above benchmark range and at the end of the assessment period, 65 students were within the benchmark range, then 74% of the students within the school were considered established (i.e., at benchmark) at the end of the assessment period. Outcome percent established index was similar in concept as measuring the schools' end of the year outcomes on a state level assessment test since initial starting points of the students are not accounted for.

Percent Adequate Progress

The second index was the percentage of students within a school making adequate progress. Adequate progress was defined as students improving their instructional recommendation or remaining at benchmark over a specified period of time. A SAS program was created that investigated the 4041-5655 schools for whom this index was computed. First, the program created a count of the number of students who scored in the intensive range, strategic range, barely benchmark range, and well above benchmark range for each school with complete data at the beginning of the sampling period. For each school, a count was created of the number of students that began in the intensive

range and reached the strategic or benchmark range (i.e., students in the intensive range that made adequate progress), the number of students that began in the strategic range and reached the benchmark range (i.e., students in the strategic range that made adequate progress), and those that began in the barely benchmark and well-above benchmark range and remained at the benchmark range at the end of the sampling period (i.e., students in the benchmark range that made adequate progress). The percentage of students that made adequate progress was computed for each school by determining the total number of students making adequate progress divided by the total number of students in a school with complete data.

For example, percent adequate progress for one school was computed as follows:

(a) 40 students scored in the intensive range at the beginning of the assessment period and 10 of those students scored in the strategic or benchmark range at the end of the assessment period; (b) 60 students scored in the strategic range at the beginning of the assessment period and 40 of those students scored in the benchmark range at the end of the assessment; and (c) 100 students who scored in the barely benchmark range and well-above benchmark at the beginning of the assessment period and 80 of those students remained at the benchmark range. Overall 130 of 200 students or 65% of the students in that school had made adequate progress.

The percent adequate progress index differed from the outcome percent established index because the percent adequate progress index included students who made adequate progress, but did not reach a benchmark recommendation level. For example, students who initially scored in the intensive range and then scored in the

strategic range at the end of the sampling period were included in the percent adequate progress index, but were not included in the outcome percent established index. Neither the percent adequate progress nor the outcome percent established index controlled for initial starting point of the students (e.g. computing an index only for the students in the intensive range) and both indices included students in the well-above benchmark range. Due to this lack of control this index would be similar in concept as measuring the schools' end of the year outcomes on a state level assessment test.

Intensive Percent Adequate Progress

The third index examined was the percentage of students that began in the intensive range and made adequate progress. Intensive adequate progress was defined as students improving their instructional recommendation from intensive to either a strategic or benchmark instructional recommendation. This index controlled for the initial starting point of the student by examining only students who began within the intensive range. A SAS program was created that investigated the 3970-5601 schools for whom this index was computed. First the program created a count of the number of students who scored in the intensive range at the beginning of the assessment period for each school. For each school, a count then was created of the number of students that initially scored in the intensive range that reached strategic or benchmark at the end of the assessment period (i.e., the students in the intensive range that made adequate progress). The percentage of students in the intensive range that made adequate progress was computed for each school by determining the total number of students in the intensive range that made

adequate progress and divided that by the total number of students who scored in the intensive range at the beginning of the assessment period.

For example, if a school had 60 students who scored in the intensive range at the beginning of the assessment period, and 40 of those students scored in the strategic or benchmark at the end of the assessment period, the percent of students in the intensive range who made adequate progress for the school would be 66%. This allowed for schools to be compared and contrasted. For example, one school in the sample could have had 20% of their students who began in the intensive range reach strategic or benchmark range, while another could have 66%. The school in which 66% made adequate progress may be more effective at addressing the needs of students in the intensive range than the school that had 20% making adequate progress. The intensive percent adequate progress index also schools to see if they are addressing the needs of the students that need intensive reading support, such as students receiving tier 3 intervention supports.

Strategic Percent Adequate Progress

The fourth index was the percentage of students that began in the strategic range and made adequate progress by reaching the benchmark range. Strategic percent adequate progress was defined as students improving their instructional recommendation from strategic to a benchmark instructional recommendation. This index controlled for the initial starting point of the student by requiring that all students began within the strategic range. A SAS program was created that investigated the 4011-5653 schools for whom this index was computed. First the program created a count of the number of students who scored in the strategic range at the beginning of the sampling period for each school.

For each school, an additional count was created of the number of students who initially scored in the strategic range that reached the benchmark range at the end of the sampling period for each school (i.e., the students in the strategic range that made adequate progress). The percentage of students that made adequate progress in the strategic range was computed for each school by determining the total number of students in the strategic range that made adequate progress and divided by the total number of students who scored in the strategic range at the beginning of the assessment period.

For example, if a school has 50 students who scored in the strategic range at the beginning of the assessment period, and 20 of those students scored in the benchmark range at the end of the assessment period, then 40% of the students in the strategic range made adequate progress in that school. A school where 80% of students in the strategic range make adequate progress would have higher rates of adequate progress for students who are in the strategic range. Again, this allows for the comparison of schools and may provide an index of the effectiveness of schools in addressing the reading needs of the students that need strategic reading support, such as students receiving tier 2 intervention supports.

Barely Benchmark Percent Adequate Progress

The fifth index was the percentage of students who initially scored in the barely benchmark range, defined as being within 20 points of benchmark, and made adequate progress by remaining at benchmark. This index controlled for the initial starting point of the student by requiring that all students began within the barely-benchmark range. A SAS program was created that investigated the 4039-5653 schools for whom this index

was computed. First the program created a count of the number of students who scored in the barely benchmark range at the beginning of the sampling period for each school. An additional count was created of the number of students who initially scored in the barely benchmark range and remained in the benchmark range at the end of the sampling period for each school (i.e., the students in the barely benchmark range that made adequate progress). The percentage of students that made adequate progress in the barely benchmark range was computed for each school by determining the total number of students in the barely benchmark that made adequate progress and divided by the total number of students who scored in the barely benchmark range at the beginning of the assessment period.

For example, if a school has 80 students who scored in the barely benchmark range at the beginning of the assessment period, and 75 of those students scored in the benchmark at the end of the assessment period, the percent of students in the barely benchmark range who made adequate progress for the school would be 94%. In contrast, a school that had only 25% of their students remain in the benchmark range would need to determine how best to meet the needs of the students that are in the barely benchmark range. Again, this allowed for schools to determine if they were addressing the reading needs in the core curriculum, such as in tier 1, of the students that have reached benchmark status, but aren't solidly at that recommendation level.

School-wide High Rates of Adequate Progress

The sixth index was school-wide high rates of adequate progress. To be considered a school with system-wide high rates of adequate progress, a school needed to

meet three criteria. Schools needed to be within the upper 1/3 of schools in the rate of students in the intensive and strategic range that made adequate progress, or have 80% of students in the intensive and strategic range in a school that made adequate progress. Also, schools needed to be within the upper 1/3 of schools in the rate of students in the barely benchmark range that made adequate progress, or have 95% of the students in the barely benchmark range in the school that made adequate progress. Initial starting point was controlled by taking into account where all students began: within either the intensive, strategic, or barely benchmark recommendation ranges.

A SAS program was created that investigated the 4018- 5547 schools for whom this index was computed. First, the program created a count of the number of students who scored in the intensive range, the students who scored in the strategic range, and the students who scored in the barely benchmark range at the beginning of the sampling period. A count was then created of the number of students that began in the intensive, strategic, and barely benchmark range that made adequate progress at the end of the sampling period. The number of students that made adequate progress was divided by the total number of students who scored in the intensive, strategic, and barely benchmark ranges at the beginning of the assessment period. This calculation was completed for each school. The top 1/3 of schools within the sample or 80% of the students in each school that had adequate progress in the intensive and strategic range and the top 1/3 or 95% of the students in the barely benchmark was calculated. Out of 4018-5547 schools for whom this index was computed, only 213-604 were schools that had high rates of adequate progress.

For example, computation of the system-wide high rates of adequate progress index for one school was as follows: (a) 40 students scored in the intensive range at the beginning of the assessment period and 32 of those students scored in the strategic or benchmark range at the end of the assessment period; (b) 60 students scored in the strategic range at the beginning of the assessment period and 48 of those students scored in the benchmark range at the end of the assessment; and (c) 100 students who scored in the barely benchmark range at the beginning of the assessment period and 95 of those students remained at the benchmark range. Overall this school had 80% of their students that scored in the intensive and strategic ranges and 95% of their students that scored in the barely benchmark range make adequate progress. This school qualified as a school with system-wide high rates of adequate progress . A school with system-wide high rates of adequate progress appears to be relatively effective a meeting the needs of students at all three tiers of instruction.

Next Steps

SAS programs were written for all the indices for the beginning of year to the middle of the year and the middle of the year to the end of the year, creating the semester-by-semester analyses. Programs were also created for beginning of the school year to the end of the school year, creating the yearly analyses. The analysis process with 6 indices and 3 time periods was repeated for each year of archival data from 2005-2006, 2006-2007, and 2007-2008. Depending on which index and research question was examined, indices were calculated for 3970-5655 schools.

Data Analysis

Correlational and Kappa analyses were conducted to address the three research questions, as described below.

1. Which indices provided the most stable information about school based adequate reading progress from first semester to second semester?

For this analysis, each index was calculated for first semester and second semester for the 2005-2006, 2006-2007, and 2007-2008 school years. The first semester and second semester for each of the six indices were correlated. This resulted in one stability coefficient for each index per year. For example, the percent established index had 3 semester-to-semester stability coefficients, one each for 2005-2006, 2006-2007, and 2007-2008. The results are displayed in a correlation table. The only exception to this was the high rates of adequate progress index, for which Kappa analyses were utilized and the results were displayed in a Kappa table. This allowed for the typical stability for each index to be investigated, as well as the variation in stability indices across years.

2. Which indices provided the most stable information from the first year, 2005-2006, to the second year, 2006-2007, to the third year, 2007-2008?

For this analysis, each of the six indices were calculated from the beginning of the year to the end of the year for each of the school years 2005-2006, 2006-2007, and 2007-2008. After the whole year indices were calculated, then the three years were correlated for each index. For example, the whole year percent established index was correlated from the first year, 2005-2006, to the second year, 2006-2007; from the second year, 2006-2007, to the third year, 2007-2008; and from the first year, 2005-2006, to the third

year, 2007-2008. Each index had 3 across year reliability coefficients. The results are displayed in a correlation table. Again, the exception to this was the high rates of adequate progress index, for which Kappa analyses were utilized and the results were displayed in a Kappa table.

3. How do the semester indices relate to the yearly indices?

For this analysis, data from the three-year sampling period were utilized. The first semester index was correlated with the whole year index, and the second semester index was correlated with the whole year index. The high rates of adequate progress index was examined similarly using the Kappa coefficient. For example, the first semester percent established index was correlated with the whole year outcome percent established index. This was completed for each year individually. The results were displayed in correlation tables or a Kappa table as warranted. The stability was then examined for each index.

Summary

To summarize, each of the six indices were examined over a three-year period. Each index was comprised of schools that had at least 40-second grade students and complete data depending on the research question. Correlation or Kappa coefficients were run for each index in each of the three research questions. Results and interpretation are presented in the following two sections.

CHAPTER IV

RESULTS

Each student in each school within the data set had complete data for a specific time period. For example, in the semester-to-semester analysis for the second semester (i.e., middle to end of the school year) a student in the outcome percent established was required to have a benchmark score in the middle of the year and at the end of the year. Treating the indices as indicators of school-wide effectiveness requires the inference that the gains that a particular student made over the semester or the year time period were attributable to the reading instruction that was being provided by the school. Depending on the particular research question and index being examined, between 3970 and 5655 schools were included in the data set. No outliers were deleted from the data set since they represented how the individual schools entered their data into the DIBELS Data System and were representative of the students in that particular school. Correlations and Kappa coefficients were used to examine the relationship among the different indices and time periods.

The stability of the indices of adequate reading progress or lack thereof may represent relative freedom from error (i.e., reliability) or may represent an accurate measure of a true state that has changed. For example, if the semester-to-semester intensive percent adequate progress index had a low stability coefficient then it is possible that the schools in the first semester were generally effective with helping their

students make adequate progress in reading and in the second semester there were fewer students who scored within the intensive instructional recommendation so they did not have as high a percentage of students that made adequate progress.

For ease of interpretation, the results are presented separately for each research question. The results are structured to first examine the outcome percent established index followed by the percent adequate progress index. Both of these indices do not limit the initial starting point of a student within a school (i.e., for the intensive percent adequate progress index the students must start the sampling period within the intensive range). The tables for the remainder of the indices are presented in the following order: intensive percent adequate progress, strategic percent adequate progress, barely benchmark percent adequate progress, and finally system-wide high rates of adequate progress. All of these indices accounted for the initial starting point of the student in some way. General observations regarding the stability of the indices are discussed in the context of the tables. Additional interpretation of implications follows in the discussion chapter.

Descriptive Statistics

Descriptive statistics for the first semester (beginning to middle of the school year), second semester (middle to end of the school year), and for the year (beginning to end of the school year) are presented in Table 1 and Table 2. Table 1 addresses the descriptive statistics for the following indices: outcome percent established and percent adequate progress. Neither of these indices accounted for the initial starting point of the

student. These indices included all students regardless of whether they initially scored within the intensive, strategic, barely-benchmark, and well-above benchmark ranges.

Table 1

Descriptive Statistics for Second Grade by Outcome between Cohort Year and Index Time Period

Cohort year	First semester			Second semester			Year		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Outcome percent established									
2005-06	4561	63.35	15.96	4570	56.46	17.18	4621	56.38	17.18
2006-07	5628	65.34	15.56	5636	57.42	16.20	5655	57.46	16.17
2007-08	5469	66.94	15.01	5474	59.54	15.68	5481	59.55	15.67
Percent adequate progress									
2005-06	4561	65.98	15.48	4570	58.71	16.79	4621	59.02	16.91
2006-07	5628	67.77	15.13	5636	59.50	15.76	5655	59.94	15.85
2007-08	5469	69.40	14.55	5474	61.57	15.21	5481	62.09	15.32

Note. All counts were of school level summaries of students with complete ORF data for the beginning of a time period to the end of a time period and 40 or more students.

The descriptive statistics for the indices that accounted for the initial starting point of the student are reported in Table 2. Interestingly, the mean percentage of students making adequate progress appears to be higher in the first semester than in the second semester for all of the indices. Between the first and second semester, there was a difference of 30 percent in students making adequate progress in the barely benchmark

range, which represents a 33% decrease from the first semester percent of barely benchmark making adequate progress. For the strategic range, there was a first to second semester difference of 44% in strategic students making adequate progress, a decrease of 62% percent from the first semester. . For the intensive range, there was a first to second semester difference of 6% in intensive students making adequate progress, a decrease of 32% percent from the first semester. Overall, the barely benchmark range had the highest percentage of students making adequate progress, followed by the strategic range and then intensive range. The mean percentage making adequate progress for the entire year mirrors the first semester adequate progress. Perhaps this is indicative of reading skills taught in second grade.

Alternatively, the number of students making adequate progress in the intensive, strategic, and barely benchmark ranges may have decreased in the second semester because the students that were close to making adequate progress reduced their risk in the first semester. The remaining students in these recommendation levels for second semester may have been scoring at the lower end of the recommendation level and were unable to make adequate progress over the semester time period. For example, students may have moved from being at the top of the intensive range to being at the bottom of the strategic range. For those students, continued adequate progress to the benchmark range in the second semester may be particularly challenging. This process may have affected the results of the indices in the research question that investigated semester-to-semester stability.

Table 2

Descriptive Statistics for Second Grade by Initial Status and Outcome between Cohort Year and Index Time Period

Cohort year	First semester			Second semester			Year		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Intensive percent adequate progress									
2005-06	4513	17.34	16.60	4533	12.53	13.07	4573	17.76	17.14
2006-07	5575	17.99	17.20	5605	11.96	12.84	5601	17.92	17.21
2007-08	5412	18.69	17.03	5435	12.25	13.37	5424	18.75	16.96
Strategic percent adequate progress									
2005-06	4557	43.14	20.43	4551	17.83	18.20	4617	31.78	21.06
2006-07	5626	43.83	19.84	5619	15.86	16.74	5653	30.23	19.17
2007-08	5467	45.88	20.03	5451	16.85	17.45	5479	32.58	19.80
Barely benchmark percent adequate progress									
2005-06	4560	89.80	12.05	4569	61.65	19.49	4620	73.53	20.73
2006-07	5626	90.68	10.64	5636	60.04	18.98	5653	73.25	19.00
2007-08	5469	91.75	9.70	5474	61.85	18.45	5481	75.61	17.69

Note. All counts were of school level summaries of students with complete ORF data for the beginning of a time period to the end of a time period and 40 or more students.

Research Question 1

To determine which indices provided the most stable information about school-based adequate reading progress from the first semester to the second semester, stability coefficients (i.e., 4-month test-retest reliability coefficients) were calculated. These were calculated from correlations or Kappa coefficients for each index for first semester and

second semester for the 2005-2006, 2006-2007, and 2007-2008 school years. For example, outcome percent established had stability coefficient for 2005-2006, 2006-2007, and 2007-2008 for school based adequate reading progress from the first semester and second semester. This allowed for an examination of the typical stability for each index to be investigated within a one-year period, as well as the variation over three years.

The first index, outcome percent established, referred to the percent of students that scored in the benchmark range at the end of the semester, regardless of where the students initially scored at the beginning of the semester. This included those students that scored in the barely benchmark range (i.e., with the first 20 points of benchmark score), as well as those that scored in the well above benchmark range (i.e., more than 20 points above the benchmark score). Outcome percent established also included students that scored in the strategic range, and that scored in the intensive range. The stability coefficients for outcome percent established are displayed in Table 3.

Across the 3 study years, the mean stability coefficient within a year (i.e., from the first semester to the second semester) was .89. The mean stability coefficient for each semester from one year to the next was .79 (i.e., first semester of one year to first semester of the next year, and second semester of one year to second semester of the next year). Across two years, the mean stability coefficient for each semester was .74.

Table 3

Outcome Percent Established: Stability Coefficients of 2nd Grade First Semester to Second Semester

Year	'05-'06 First semester	'05-'06 Second semester	'06-'07 First semester	'06-'07 Second semester	'07-'08 First semester	'07-'08 Second semester
2005-2006						
First semester	—	—	—	—	—	—
2005-2006						
Second semester	0.90	—	—	—	—	—
2006-2007						
First semester	0.80	0.78	—	—	—	—
2006-2007						
Second semester	0.74	0.78	0.89	—	—	—
2007-2008						
First semester	0.76	0.74	0.80	0.75	—	—
2007-2008						
Second semester	0.70	0.72	0.74	0.78	0.89	—

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4039$ to 5636 schools with 40 or more students and complete 2nd grade ORF data.

Thus, percent established was highly stable as an index of school context, however the stability may have been due to the inclusion of students in the well above benchmark range along with the intensive, strategic, and barely benchmark ranges. Consequently, outcome percent established may be, to some extent, an index of status of the school or community and, as such, may be more likely to be stable. It is probable that the schools that began the semester with many students at benchmark or well above

benchmark, such as those in schools that have high achieving students or are in communities are typically high achieving, are likely to have many of their students remain at benchmark or well above benchmark at the end of the semester. It is also unlikely that students who score initially in the well above benchmark range at the start of the semester would score in the strategic range at the end of the semester. Thus, the students who score within the well above benchmark range may tend to inflate the stability of the outcome percent established index for a school.

The second index, percent adequate progress, refers to the percent of students that move from scoring within one instructional recommendation to a higher instructional recommendation (i.e., move from being at higher risk to being at reduced risk) or remain within the benchmark range for a particular school. For example, if a student moved from scoring within the intensive range to scoring within the strategic range from the fall benchmark to the winter benchmark, then that student would have made adequate progress for the first semester. The stability coefficients for percent adequate progress are presented in Table 4. The percent adequate progress index also included students whose scores started out either at an intensive, strategic, or benchmark (including barely benchmark or well above benchmark) range.

Overall, the mean stability coefficient within a year was .85 and across one year the mean stability coefficient was .78 and across 2 years it was about .72. This indicated that the percent adequate progress index provided a highly stable measure of school based adequate reading progress within and across semesters within the three year time period.

Table 4

Second Grade, Semester-to-semester Stability Coefficients for Percent Adequate Progress

Year	'05-'06 First semester	'05-'06 Second semester	'06-'07 First semester	'06-'07 Second semester	'07-'08 First semester	'07-'08 Second semester
2005-2006						
First semester	—	—	—	—	—	—
2005-2006						
Second semester	0.86	—	—	—	—	—
2006-2007						
First semester	0.79	0.76	—	—	—	—
2006-2007						
Second semester	0.71	0.76	0.85	—	—	—
2007-2008						
First semester	0.74	0.72	0.79	0.72	—	—
2007-2008						
Second semester	0.67	0.70	0.70	0.76	0.85	—

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4041$ to 5628 schools with 40 or more students and complete 2nd grade ORF data.

However, as with the outcome percent established index, if a particular school had a large number of students that scored in the well-above benchmark range and those students remained at least in the benchmark range at the end of the semester those students would inflate the stability of the percent adequate progress index for that school. An inflated measure of school effectiveness could become problematic if the school has a number of students scoring in the well-above benchmark range and a number of students

scoring in the strategic or intensive ranges because it is possible that the school could not be meeting the reading needs of the students scoring in the strategic and intensive ranges, but may have a relatively high index of percent adequate progress due to the students scoring in the well-above benchmark range.

The indices that follow, intensive percent adequate progress, strategic percent adequate progress, and barely-benchmark percent adequate progress, help to address the inflation that the students that score in the well-above benchmark range have on the stability of outcome percent established and percent adequate progress indices by controlling for the initial starting point of students within the school. By investigating selected subsets of students within a school, these indices may more accurately reflect the instruction that is being provided to students that need different levels of support. To begin, Table 5 represents the index of the intensive percent adequate progress, which focused on the students within schools with the most intensive reading needs.

Intensive percent adequate progress referred to the percent of students who scored in the intensive range at the start of the semester that scored in either the strategic range or benchmark range at the end of a semester. Intensive percent adequate progress was an index based only on students that started out in the intensive range at the start of the semester. For students starting the semester scoring in the intensive range, adequate progress entails ending the semester scoring in either the strategic range or benchmark range. The intensive percent adequate process index was conceptualized as an indication of the effectiveness of the school's third tier of reading support within an RTI context in

that students scoring in the intensive range are likely to be receiving intensive and individualized, tier 3 support.

Table 5

Second Grade, Semester-to-semester Stability Coefficients for Intensive Percent that Make Adequate Progress.

Year	'05-'06 First semester	'05-'06 Second semester	'06-'07 First semester	'06-'07 Second semester	'07-'08 First semester	'07-'08 Second semester
2005-2006						
First semester	—	—	—	—	—	—
2005-2006						
Second semester	0.13	—	—	—	—	—
2006-2007						
First semester	0.24	0.13	—	—	—	—
2006-2007						
Second semester	0.03	0.16	0.13	—	—	—
2007-2008						
First semester	0.23	0.11	0.26	0.08	—	—
2007-2008						
Second semester	0.06	0.13	0.06	0.14	0.10	—

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 3970$ to 5560 schools with 40 or more students and complete 2nd grade ORF data.

Overall, the mean stability coefficient within a year was .12 and the across one year was .20 and across 2 years was about .18. This indicated that the intensive percent adequate progress index was not a stable measure of school-based effectiveness within and across semesters within the three year time period. Interestingly, the stability the first

semester of year one compared to the first semester of year two and year three was stronger than that of first semester to the second semester within a year. After examining the differences in mean percentages of students making adequate progress from the first semester to the second semester within the same year, it may be possible that there were more students in the intensive range that were on the cusp of the strategic range in the first semester and therefore made adequate progress. In contrast, in the second semester there may have been fewer overall numbers of students scoring in the intensive range (and thus they may have represented more challenging needs) or perhaps more students that were scoring on the lower end of the intensive range and they were difficult to support to make adequate progress. The potential difference in the sample size or the initial skill level within the intensive range may account for the lack of stability across the semesters.

Making adequate progress for a student scoring in the intensive range was a very rigorous standard. These were the students that were scoring below 26 words read correctly per minute in the fall and below 52 words read correctly in the winter. In order for them to make adequate progress, their reading skills would have to improve to from below 26 to above 52 words correct per minute in the first semester, or from below 52 to above 69 words correct per minute in the second semester. It is possible that there may have been students that were only reading a few words per minute who still made substantial progress. Instruction for the students that were scoring closer to 0 words read correctly per minute would look different than the students scoring closer to the strategic range. If a student who scored 5 words read correctly per minute in the fall progress such

that, by the winter benchmark they read 36 words correctly per minute, that student would have made substantial reading progress, however in the intensive percent adequate progress they would not be counted as making adequate progress because they did not reach the strategic recommendation of 52 words read correctly per minute by the middle of the year.

The stability coefficients of the strategic percent adequate progress index, which focused on the students within schools with strategic reading needs, are reported in Table 6. Strategic percent adequate progress referred to the percent of students who scored in the strategic range at the start of the semester and who scored in the benchmark range at the end of a semester. Strategic percent adequate progress was an index based only on students that scored in the strategic range at the start of the semester. For students that scored at the beginning of the semester in the strategic range, adequate progress entailed ending the semester with scores in the benchmark range. The strategic percent adequate progress index was conceptualized as an indication of the school-based effectiveness of second tier of reading support within an RTI context since typically students that score within the strategic range on DIBELS receive tier 2 intervention support within schools.

Overall, the mean stability coefficient within a year was .11, across one year was .33, and across two years was .27. The strategic percent adequate progress index was also not a stable measure of school-based rates of adequate reading progress within and across semesters for the three year time period. Interestingly, the stability of the one year and two year measures for the first semester was again stronger, roughly double, that of first semester to second semester stability.

Table 6

Second Grade, Semester-to-semester Stability Coefficients for Strategic Percent that Make Adequate Progress.

Year	'05-'06 First semester	'05-'06 Second semester	'06-'07 First semester	'06-'07 Second semester	'07-'08 First semester	'07-'08 Second semester
2005-2006						
First semester	—	—	—	—	—	—
2005-2006						
Second semester	0.15	—	—	—	—	—
2006-2007						
First semester	0.40	0.21	—	—	—	—
2006-2007						
Second semester	0.07	0.24	0.09	—	—	—
2007-2008						
First semester	0.36	0.20	0.45	0.10	—	—
2007-2008						
Second semester	0.07	0.17	0.09	0.23	0.10	—

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4011$ to 5610 schools with 40 or more students and complete 2nd grade ORF data.

The lack of stability between the two semesters may be attributed to higher rates of adequate progress in the first semester versus that of the second semester. While the students that score in the strategic range have higher reading skills than those that score in the intensive, it is possible that the students that were scoring within the low end of the strategic range could have seen large amounts of growth over a semester time period, but were unable to make enough growth to enter the benchmark range. There may be more

students at the low end of the strategic range during second semester than during first semester.

The stability of the index of barely benchmark percent adequate progress is reported in Table 7. Barely benchmark percent adequate progress focused on the students within schools that scored within the benchmark recommendation status at the beginning of the semester, but because they were on the lower end of the benchmark range they may have had more difficulty maintaining scores in the benchmark range over a semester period. Barely benchmark percent adequate progress referred to the percent of students who scored within the first 20 points of the benchmark range and who remained scoring in the benchmark range over a semester period. Barely benchmark adequate progress was an index based only on students that started out scoring in the barely benchmark range at the start of the semester. For students starting the semester in the barely benchmark range, adequate progress entails ending the semester with scores within the benchmark range. It was conceptualized as an indication of school-based effectiveness of first tier of reading support within an RTI context since students whose scores fell within the barely benchmark range likely were receiving only core or Tier 1 reading instruction.

Overall, the mean stability coefficient within a year was .13, across one year was .30, and across two years was about .18. Barely benchmark adequate progress also was not a stable measure of school based adequate reading progress within and across semesters within the three year time period. This remained similar to the other indices that controlled for initial starting point.

Table 7

Second Grade, Semester-to-semester Stability Coefficients for Barely Benchmark Percent that Make Adequate Progress.

Year	'05-'06 First semester	'05-'06 Second semester	'06-'07 First semester	'06-'07 Second semester	'07-'08 First semester	'07-'08 Second semester
2005-2006						
First semester	—	—	—	—	—	—
2005-2006						
Second semester	0.17	—	—	—	—	—
2006-2007						
First semester	0.29	0.19	—	—	—	—
2006-2007						
Second semester	0.09	0.24	0.13	—	—	—
2007-2008						
First semester	0.21	0.15	0.30	0.10	—	—
2007-2008						
Second semester	0.05	0.15	0.10	0.35	0.09	—

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4039$ to 5636 schools with 40 or more students and complete 2nd grade ORF data.

Two constants with the previous indices that controlled for initial starting point, strategic percent adequate progress and the intensive percent adequate progress, were that there remained a drop in stability from the first semester to the second semester and all the indices were about equivalently stable. Again, if in the first semester there were a number of students on the cusp of scoring within the well-above benchmark range and their scores increased to more than twenty points above benchmark they would not be

examined in the next semester to see if they continued to make adequate progress. This may impact the number of students in the barely benchmark range the next semester and therefore may impact the number of students that made adequate progress.

A main difference between the barely benchmark percent adequate progress index and the other two indices that accounted for initial starting point is as follows. If a student is at the low end of the benchmark range at the start of the semester and then makes enough growth to remain in the low end of the benchmark range at the end of the semester, then the student would have still made adequate progress. In contrast, with the intensive and strategic percent adequate progress indices, if the student's scores only increased enough to keep his or her instructional recommendation the same (i.e., a student was in the strategic range at the beginning of the semester and remained in the strategic range at the end of the semester), then he or she would not have made adequate progress. This could potentially affect the stability of the barely benchmark percent adequate progress index compared to the intensive and strategic percent adequate progress indices because the students scoring in the barely benchmark range may have an easier time making adequate progress and therefore be larger in magnitude. However, this does not appear to be the case based on the results from the semester-to-semester analysis.

The indices just reported all took into consideration the starting point of one particular subset of students within a school: intensive percent adequate progress, strategic percent adequate progress, and barely benchmark percent adequate progress. Overall, none these indices provided very stable measures of school-wide rates adequate

reading progress. Two potential interpretations of these findings are that (a) the indices may not be stable in terms of reliability from semester to semester, or (b) the indices may accurately measure the rates of progress of those students within those instructional recommendations and the semesters may have different in rates of adequate progress in actuality. For example, if a group of students in the intensive range within a school made adequate progress for first semester, then there may be a smaller number of students left in the intensive range and a larger number of students at the low end of the strategic range for the second semester. The smaller number of students remaining in the intensive range may have a qualitatively different set of education needs. The students scoring in the intensive and strategic ranges at the beginning of the second semester might make substantial progress, but based on their initial starting point their progress may not be sufficient to move up an instructional recommendation. As a result, these students would not have made adequate progress and the correlation between the first semester index and the second semester index would be low. The school may be meeting the needs of the majority of the students that are scoring within the intensive range, but the stability coefficient may not accurately reflect that effectiveness.

Another possibility is that there is an actual difference in the effectiveness of the school from semester to semester. For example, there may be a change in the focus of instruction from the first semester to the second, with the first semester of second grade emphasizing word reading and decoding and advanced phonics skills, and in the second semester the school may emphasize reading fluently or comprehension strategies. The

change in emphasis may meet the needs of some groups of students but not others, or the school may be differentially effective in providing instruction for each emphasis.

However, what if a school has been actively and adequately supporting all of their students at their instructional level (i.e., students scoring in the intensive, strategic, and barely benchmark ranges) and was interested in an overall index or standard that would give them an indication that all levels of students were making high rates of adequate progress? These schools are not the typical schools in the United States, instead they are the exceptional schools that are able to provide reading instruction that was generally effective for the majority of their student population and sustain that progress over a period of time. System-wide high rates of adequate progress is the index that examines this construct.

To be considered a school with system-wide high rates of adequate progress, a school needed to meet several criteria. Schools needed to be within the upper 1/3 of schools for the sample in the proportion of students in the intensive and strategic range that made adequate progress, or have 80% of students in the intensive and strategic range in a school that made adequate progress (whichever is lower). Also, schools needed to be within the upper 1/3 of schools for the sample in the rate of students in the barely benchmark range that made adequate progress, or have 95% of the students in the barely benchmark range in the school that made adequate progress (whichever is lower). Schools meeting all of these criteria were considered to be schools that had system-wide high rates of adequate progress at all instructional levels. The Kappa coefficients for system-wide high rates of adequate progress are displayed in Table 8.

Table 8

Second Grade, Semester-to-semester Stability of Decisions with Schools that have System Wide High Rates of Adequate Progress

Cohort	Number of schools	Number of school with system-wide high rates of adequate progress			Kappa
		First semester	Second semester	Both semesters	
2005-2006	4486	451	375	65	0.03
2006-2007	5547	552	467	69	0.02
2007-2008	5380	604	400	74	0.03

Note. All Kappa's were significant. Kappa's were based on $n = 4486$ to 5380 schools with 40 or more students, complete 2nd grade ORF data, and system-wide high rates of adequate progress.

Kappa has a clear interpretation, which is percent improvement over chance agreement. According to Viera and Garrett (2005) the intent of Kappa is to allow for a quantitative measure of the magnitude of agreement. It measures the amount of precision, or reliability, in the agreement. However, there are no clear evaluative guidelines for Kappa. Landis and Koch (1977) suggested that the scale ranged as the following: less than 0.00 as poor agreement, 0.00 to 0.20 as slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1.00 as almost perfect agreement. Based on an evaluation of the Kappa coefficients, there was very little stability, just a slight increase in agreement beyond chance, between each semester. However, this doesn't necessarily mean that this index of system-wide high rates of adequate progress wasn't an accurate measure of schools that

have system-wide high rates of adequate progress. It just indicated that the schools that have system-wide rates of adequate progress were not plentiful and typically weren't able to maintain the level of progress over consecutive semesters. Perhaps the semester time period may be too unstable and a longer period of time, such as a year, may provide an increase in stability for system-wide high rates of adequate progress.

Research Question 2

In the previous research question year-to-year was examined by comparing the same semester over a period of years (i.e., first semester of 2005-2006 year to the first semester of 2006-2007). Instead, this research question examines a full school year (i.e., from the beginning of the year to the end of the year) and then inspects the stability of each index over a period of one whole year or two whole years. To determine which indices provided the most stable information about school based adequate reading progress from the first year, 2005-2006, to the second year 2006-2007, to the third year, 2007-2008, stability coefficients (i.e., reliability coefficients) were calculated using correlations or Kappa coefficients. Stability coefficients were calculated from the beginning of the year to the end of the year and the three years were compared for each index. For example, outcome percent established was correlated from the first year, 2005-2006, to the second year, 2006-2007; from the second year, 2006-2007, to the third year, 2007-2008; and from the first year, 2005-2006, to the third year, 2007-2008. This allowed for the typical stability for each index to be investigated over a three-year period. Table 9 examines the year-to-year stability coefficients for all of the indices with the

exception of kappa coefficients for the year-to-year system-wide high rates of adequate progress.

Table 9

Second Grade, Year-to-year Stability Coefficients by Index

Index	Year-to-year stability		
	2005-06 to 2006-07	2005-06 to 2007-08	2006-07 to 2007-08
School-wide, all-student indices			
Outcome percent established	0.78	0.72	0.78
Percent adequate progress	0.76	0.70	0.76
Indices given initial recommendation level			
Intensive percent adequate progress	0.26	0.22	0.25
Strategic percent adequate progress	0.44	0.38	0.46
Barely Benchmark percent adequate progress	0.40	0.32	0.44

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4019$ to 5655 schools with 40 or more students and complete 2nd grade ORF data.

Overall, there were a number of similarities between the year-to-year results and the semester-to-semester results. The school-wide, all-student indices tended to have a higher level of stability than the indices that controlled for the initial recommendation level. For the outcome percent established index, the mean stability coefficient across one year was an average of .78 and across two years was .72. With the percent adequate progress index, the mean stability coefficient across one year was .76 and across two

years was .70. However, as in the previous research question, the increased stability may have been inflated due to the students that scored in the well-above benchmark range for both outcome percent established and percent adequate progress indices.

With the indices that controlled for initial starting point, the year-to-year indices tended to display somewhat higher levels of stability than within the semester-to-semester indices (i.e., first semester to second semester). For the intensive percent adequate progress index, the mean stability coefficient across one year was .26 and across two years the mean stability was .22. Again, intensive percent adequate progress was not a stable measure of school-wide rates of adequate reading progress. In contrast to the semester-to-semester analysis, the lack of stability isn't likely due to declining numbers of students scoring within the intensive recommendation over the three-year period. It is likely that schools would have similar numbers of students scoring within the intensive range at the beginning of second grade over the three-year period. Similar to the semester-to-semester indices, the lower stability could indicate that the students that scored within intensive range made progress, but were unable to reach a strategic recommendation at the end of the year. Perhaps in second grade, students that scored in the low end of the intensive range may take more than one year to make adequate progress with intensive interventions.

For the strategic percent adequate progress index, the mean stability coefficient across one year was an average of .45 and across two years was .38. In addition, the barely benchmark percent adequate progress index had a mean stability coefficient across one year an average of .42 and across two years an average of .32. This indicated that

both of these indices were moderately stable measures of school-wide rates of adequate reading progress across a three year time period. Both of these stability coefficients were higher than coefficients for the intensive percent adequate progress index, which may indicate that schools were generally more stable in their effectiveness at supporting their students in the strategic and barely-benchmark ranges than those in the intensive range. As seen in the previous analysis, the year-to-year indices were more stable than the semester-to-semester indices when comparing the same index.

The general trend was that year-to-year indices tended to be more stable than the within year, semester-to-semester indices. Perhaps the longer time period of a full year allowed the schools to more intensively support their students' reading needs since the year-to-year indices tended to be more stable. If that is the case, then would the System Wide High Rates of Adequate Progress index, as seen in Table 10, also reflect stronger levels of agreement? This would indicate that schools were able to show high rates of adequate progress of all of their students that may need additional support in the area of reading (i.e., students scoring in the intensive, strategic, and barely benchmark ranges) over a period of years.

Based on the Kappa coefficients there was fair stability, or fair agreement, between years. As in the prior research question, the fair stability doesn't necessarily mean that it wasn't an accurate measure of schools that have system-wide high rates of adequate progress. It indicated that the schools that have system-wide rates of adequate progress were not plentiful, however about a third of schools were able to maintain the

level of progress for more than a year. About two-thirds of the schools did not maintain the progress over a period of more than a year.

Table 10

Second Grade, Year-to-year Stability of Decisions with Schools that have System Wide High Rates of Adequate Progress

Cohort	Number of schools	Number of school with system-wide high rates of adequate progress			Kappa
		First year	Second year	Both years	
2005-06 to 2006-07	4315	354	350	159	0.21
2005-06 to 2007-08	4018	345	318	143	0.29
2006-07 to 2007-08	5042	360	335	203	0.21

Note. All Kappa's were significant. Kappa's were based on $n = 4018$ to 5042 schools with 40 or more students, complete 2nd grade ORF data, and system-wide high rates of adequate progress.

Interestingly, the number of schools that had system wide adequate progress over a period of more than a year remained similar over the three-year period. This was also seen in the semester-to-semester analysis, but the numbers of schools that were able to maintain high rates of adequate progress were almost triple in the year-to-year analysis. Perhaps this is evidence that supporting students of all skills levels within a school achieve high rates of adequate progress requires more time than a semester's worth of instruction to make improvements.

Research Question 3

To determine how the semester indices related to the yearly indices, data from the three-year sampling period was utilized. Both the semesters (i.e., beginning of year to middle of year and middle of year to end of year) were correlated to the entire school year (beginning of year to end of year) for each index, with the exception of the system-wide high rates of adequate progress index for which Kappa coefficients were completed. For example, the outcome percent established index was correlated from the first semester to the entire year and from second semester to the entire year. This was completed for each year individually. The results were displayed in correlation tables or a Kappa table as warranted. The correlations and Kappa coefficients for the relationship between the semesters and yearly indices are displayed in Tables 11 and 12, respectively.

Overall, the school-wide, all student indices, including outcome percent established and percent adequate progress had high correlations between first and second semester indices and the whole-year indices. For the outcome percent established index, the average amount of the shared variance between the whole year index and first semester was 79%, whereas the whole year index to the second semester was 100%. Similarly, with the percent adequate progress index, the average shared variance between the whole year index and the first semester was 76% and the whole year index and the second semester was 98%. The high correlations and the large amount of shared variance indicated that the whole year indices and semester-to-semester indices are highly related. Likely this is because the same schools that are within the semester-to-semester and year-to-year outcome percent established and percent adequate progress indices. Since these

indices are based on the same schools, this may be further evidence consistent with the hypothesis that the students that scored in the well-above benchmark range are inflating the stability of outcome percent established and percent adequate progress indices.

With the indices that control for initial recommendation level, intensive percent adequate progress, strategic percent adequate progress, and barely benchmark adequate progress, there were moderate correlations among the semester indices and the whole year indices. The correlations remained similar over the three-year time period. For intensive percent adequate progress the average amount of shared variance between the whole year index and first semester was 44% and the whole year index to the second semester was 30%. The average amount of shared variance between the whole year index and first semester for strategic percent adequate progress was 41% and between the whole year index and the second semester was 33%. For the barely benchmark percent adequate progress index, the average amount of shared variance between the whole year index and the first semester was 33% and between the whole year index and the second semester was 40%. The schools that make up the year-to-year indices and the semester-to-semester indices overlap.

Based on the Kappa coefficients there was fair agreement when comparing the semesters to years. There was a fair amount of overlap between the schools that comprised the year-to-year and the semester-to-semester school-wide high rates of adequate progress. As with the indices that controlled for initial starting point, there were likely qualitative differences in schools that had school-wide high rates of adequate progress over only a year period versus only a semester period.

Table 11

Second Grade, Semester-to-year Correlation Coefficients by Index

School-wide effectiveness Index	Year		
	2005-06 to 2006-07	2005-06 to 2007-08	2006-07 to 2007-08
School-wide, all-student indices			
Outcome percent established			
First semester	0.90	0.89	0.89
Second semester	1.00	1.00	1.00
Percent adequate progress			
First semester	0.88	0.87	0.87
Second semester	0.99	0.99	0.99
Indices given initial recommendation level			
Intensive percent adequate progress			
First semester	0.67	0.66	0.66
Second semester	0.57	0.54	0.54
Strategic percent adequate progress			
First semester	0.66	0.63	0.63
Second semester	0.61	0.53	0.56
Barely benchmark percent adequate progress			
First semester	0.60	0.57	0.54
Second semester	0.64	0.64	0.62

Note. All correlations were significant at $p < .0001$. ORF correlations were based on $n = 4498$ to 5655 schools with 40 or more students and complete 2nd grade ORF data.

Table 12

Second Grade, Year-to-Semester Stability of Decisions with Schools that Have System Wide High Rates of Adequate Progress

Time period	Number of schools	Number of school with system-wide high rates of adequate progress			Kappa
		Year	Semester	Both	
2005-2006 School year					
Year to first semester	4511	330	311	213	0.32
Year to second semester	4486	290	195	245	0.44
2006-2007 School year					
Year to first semester	5572	350	371	264	0.35
Year to second semester	5547	356	286	250	0.37
2007-2008 School year					
Year to first semester	5411	586	403	288	0.26
Year to second semester	5380	573	187	287	0.36

Note. All Kappa's were significant. Kappa's were based on $n = 4486$ to 5572 schools with 40 or more students, complete 2nd grade ORF data, and system-wide high rates of adequate progress.

Summary of Results

Overall, the indices that were school-wide and included all students were the most stable. However, the stability was likely inflated due to the students within the well-

above benchmark range. The year-to-year indices, specifically those that controlled for initial skills, also tended to be more stable than the semester-to-semester indices. There are two potential explanations as to why the year-to-year indices exhibited more stability than the semester indices. First, perhaps the year time period allowed schools additional time to support interventions for their students or maturation occurred during the time period. Second, the year-to-year time period may also be less rigorous than the semester-to-semester time period. For example, it is likely that a student would have to make adequate progress in both semesters, moving from intensive to benchmark in one year, instead of making adequate progress over a year time period, moving from intensive to strategic in one year. The year-to-year indices that controlled for initial skills were moderately correlated to the same semester-to-semester indices potentially indicating there may be qualitative differences between the schools during each of the time periods.

Perhaps the most surprising result was the number of schools with school-wide high rates of adequate progress over more than a one-year period. The 143 to 203 schools that maintained school-wide high rates of adequate progress over a period of more than one year indicated that schools were able to be generally effective in helping their students achieve high rates of adequate reading progress and that schools were able to maintain that support to consecutive years of second grade students. Though the number of schools is relatively small compared to the 4018 to 5547 school sample size, it is still optimistic that schools were able to effectively support all of their students in the area of reading.

CHAPTER V

DISCUSSION

The present study examined six possible indices characterizing school-wide adequate reading progress. The goal of the study was to provide a starting point in the literature for the examination of practical, applied indices that could assist schools in determining if the instruction that they are providing in reading is effective in meeting the needs of all the students in the school. Specifically, the indices were conceptualized as a snapshot to determine if the reading instruction, as part of a three-tiered system of support, is effective within schools. An effective three-tiered system of support is a critical foundation of RTI models and indices of effectiveness are critical to provide evidence of generally effective instruction within an RTI context. This study is merely a first step in a possible line of future research on indices of school-wide rates of adequate reading progress.

In order for schools to determine if they have effective instructional strategies, programs, approaches, or techniques to support their students, an evaluative measure, such as the six proposed indices, must be created. However, prior to using such a measure the reliability, validity, and meaningfulness of the measure must be evaluated. The present study only addresses the reliability, in the form of stability, of the six proposed indices. Before a determination can be made if these indices can be used in schools, the reliability, validity, and meaningfulness of the indices must be addressed.

Stability of Indices

The stability of the six indices varied greatly depending on the index being examined, the time period, and which students' scores comprised the index. The two indices that had the greatest relative degree of stability were the outcome percent established index and the percent adequate progress index. However, since both of these indices included all students within schools in the second grade with complete data it is likely they were confounded with school wide initial skill levels. For example, if a school begins the year with many of their students scoring in the well-above benchmark range, then those students' scores will likely remain in the well-above benchmark range at the end of the semester or end of the year, thereby affecting the stability of outcome percent established and percent adequate progress. The closest comparison of these two indices to an existing construct in the literature that could be considered an index is the end of the year outcomes on state level assessments. All of them contain the drawback that initial skills within a school strongly predict outcomes.

In an attempt to overcome that drawback, indices that control for initial skills were developed. When initial skills were controlled, the year-to-year stability was greater than the semester-to-semester stability for all of the indices. The semester-to-semester indices of intensive, strategic, and barely benchmark percent adequate progress were not stable measures. Perhaps this was in large part due to higher rates of adequate progress in the first semester versus the second semester. Evidence for this can be seen in the descriptive statistics for the indices that controlled for initial starting point. The mean for intensive percent adequate progress, strategic percent adequate progress, and barely-

benchmark percent adequate progress were higher in the first semester than the second semester. The barely-benchmark percent adequate progress and intensive percent adequate progress both had about a 30% decrease from the first semester to second semester. The strategic percent adequate progress had a 62% decrease from first semester to second semester.

One possibility for the lack of stability with the semester-to-semester indices, is that schools may have been more effective in supporting adequate reading progress in the first semester and therefore the students' scores may have been starting at a lower point in the new instructional recommendation level and were not able to make adequate progress for second semester. For example, if a student made adequate progress from the intensive range to the strategic range in first semester, it is likely that he or she would have a score in the low end of the strategic range starting second semester. In order to make adequate progress in the second semester this student would have to have a score in the benchmark range by the end of second semester.

In contrast, the mean percentages of students within schools making adequate progress for the first semester mirrored that of the entire year. This could be due in part to the requirement for one semester worth of adequate progress and one year of adequate progress was identical. For instance, in the semester-to-semester example a student would have to make adequate progress from intensive to strategic in the first semester and then make adequate progress from strategic to benchmark in the second semester. When adequate progress is examined over a year period, the student would have only had to have made adequate progress from intensive to strategic over the entire school year. One

possibility for the higher stability in the year-to-year indices could be that with the longer time period for schools to help their students achieve adequate progress a less rigorous standard was created.

There also may have been a change in the focus of instruction from the first semester to the second, with the first semester of second grade emphasizing word reading and decoding and advanced phonics skills, and in the second semester the school may emphasize reading fluently or comprehension strategies. The change in emphasis may meet the needs of some groups of students but not others, or the school may be differentially effective in providing instruction for each emphasis. This could account for the higher stability in the year-to-year indices over the semester-to-semester indices. It is possible that within the year-to-year indices the students made the majority of gains in the first half of the year, which would mirror the results of the semester-to-semester analysis.

Within the specific indices that controlled for initial skill level over a year-to-year comparison, the strategic percent adequate progress index appeared to be more stable than the intensive percent making adequate progress. It is possible that students needing strategic level support are more responsive to interventions than students needing intensive levels of support. Another possible explanation is that strategic level, or tier 2, support may be provided in larger group settings with established, standard protocol, interventions so that schools may have had more resources available for strategic support. Students requiring intensive, tier 3, interventions likely needed more resources from the school to plan and implement the support. An interesting result is that the stability of the

barely-benchmark and strategic percent adequate progress indices were roughly equal.

This may be due to similar instructional support provided to these students.

One important point to make is that the lack of stability of the indices that controlled for initial starting point does not necessarily mean lack of accuracy. With these analyses only the stability was being examined. For instance, the stability coefficients indicate how well the indices measure adequate reading progress over the semester-to-semester or the year-to-year time interval. The stability of these indices does not necessarily indicate whether the indices are accurately measuring school-wide adequate reading progress, the stability only indicates if it stays the same. Within a school setting, it may be that the percentage of students making adequate progress is going to fluctuate based on the initial starting point of the students and the instruction that the students are receiving.

Overall, very few schools in the sample were able to demonstrate system-wide high rates of adequate progress and support all instructional levels of students to make adequate progress. However, from semester-to-semester and year-to-year there are schools that were able to accomplish school-wide high rates of adequate progress. The year-to-year index tended to be more stable than the semester-to-semester index. In addition, the number of schools in the year-to-year index demonstrated system-wide high rates of adequate progress were almost triple that of the semester-to-semester index. Perhaps this is evidence that supporting students of all skills levels within a school to achieve high rates of adequate progress requires more time than a semester's worth of instruction.

Practical Implications for Three Tiered Models

The most apparent practical implication from the present study would be to integrate these indices of school-based adequate reading progress into three-tiered decision models, like RTI, once additional research has been conducted regarding the accuracy of the indices. Within the three-tiered system, the indices that fit the closest with each of the tiers were as follows: barely-benchmark percent adequate progress was conceptualized as an evaluative piece of tier 1 reading instruction, strategic percent adequate progress was conceptualized as an evaluative piece of tier 2 reading instruction, and intensive percent adequate progress was conceptualized as an evaluative piece of tier 3 reading instruction. The system-wide high rates of adequate progress index was conceptualized as an evaluative piece to determine if schools were effectively supporting all three tiers of reading instruction. Outcome percent established and percent adequate progress indices were conceptualized as an indication of overall school performance. Ideally, these indices would help to answer the question of if a student in a particular school was receiving generally effective reading instruction.

Differences in stability existed between the indices that control for initial skills and those that did not. The outcome percent established and percent adequate progress indices had higher levels of stability, but the initial skill level was not controlled. Since the initial skill level wasn't controlled, it would be difficult to use these indices to evaluate the different tiers of instruction within school. It may be possible to use the outcome percent established and percent adequate progress indices as an indication of

overall school performance, but it is likely that additional factors, such as the number of students in the well above benchmark range, affected the stability.

With the indices that controlled for initial skills, the stability ranged from low to moderate depending on whether the indices are examined from semester-to-semester or year-to-year. While the indices of intensive percent adequate progress, strategic percent adequate progress, and barely-benchmark percent adequate progress translate well to specific tiers in an multi-tiered system, the lower levels of stability make it difficult to interpret these indices on a school wide level without additional research. In addition, the lower number of schools that were able to achieve system-wide high rates of adequate progress indicate that it may be too rigorous of a standard for schools to achieve and maintain over multiple semesters or multiple years. However, if a student is part of a school that demonstrates system-wide high rates of adequate progress, it may be a safe assumption that the particular student is receiving generally effective reading instruction. Additional data would need to be gathered, such as individual progress of students, before making this inference. Possible alternatives that involve less rigorous standards are addressed in the future research section.

Limitations

The results must be interpreted within the limitations of the study. The six indices that were discussed were not meant to be an exhaustive method to evaluate school-wide adequate reading progress. Since the literature base on indices of adequate reading progress is virtually non-existent, there may be many other ways to conceptualize indices that provide a quick and efficient snapshot for schools to examine whether they are

supporting their students effectively. Alternative indices utilizing HLM models and mixture models may provide a more accurate representation of school-wide adequate reading progress. More sophisticated methods of modeling the indices may be able to account for the nested nature that occurs in schools. Students are not randomly assigned into schools and it is likely that there are confounding effects that may affect the stability of the indices.

The participants in the DIBELS Data System (DDS) may entail selection effects. It is possible that schools and districts that utilize the services of the DDS may look markedly different from those that don't. The degree to which the schools utilize the features of the DDS could differ. For instance, one school may enter their DIBELS data into the DDS, but that is the extent to which they use the system. If they aren't running reports and making educational decisions that impact the instruction of their students, then it is plausible they may have different rates of adequate progress than schools that are using the students' scores to impact the instruction.

After schools collect the data, even if the schools are using the DDS and the features of the database, there is no guarantee they are using the formative assessments to guide their instruction. These schools may not have a school-wide system in place to provide differentiated levels of instruction. Even if schools were collecting DIBELS data at every benchmarking period they may not be intervening and providing their students with intervention support based on their DIBELS recommendation levels. If that were the case then it would be unlikely that adequate progress would be seen in the scores of the students, especially those that initially scored in the intensive or strategic ranges.

With the data sampling procedures, schools were required to have complete data for their students over specified time period, at least a year, depending on the research question. This required the school to have a consistent commitment to utilizing DIBELS and using the DDS to input their data. This could pose a limitation in that these schools may be inherently different than schools that don't use the DDS at all or don't use the DDS consistently. It is possible that there is a higher level of commitment for instructionally supporting students to make school-wide adequate reading progress with the schools that utilize the DDS system.

The quality of data collection of the DIBELS data that is inputted into the DDS may differ from school to school. The participating schools are responsible for the training in the administration of measures, fidelity of implementation, and the data entry. Some schools may opt to have a team come into the classroom and do the assessments with all the students, whereas other schools may opt to have teachers collect the data. There may also be differences in the level of intervention if teachers collect their own data versus having a team collect the data. Schools may also have opted to administer only one benchmark passage to each student instead of the required three passages disregarding the standardized administration of DIBELS. Although it is a limitation of the study, the way in which schools input data and administer DIBELS measures to their students is reflective of what schools actually do rather than what they could do with optimal training and support. If anything the stability that is reflected with these indices that study examines is under "real world" conditions.

Future Research

The current study provides directions for future research within the field for different indices of school-wide adequate reading progress. Since the present study evaluated only the stability of the six indices one of the next steps would be to address the accuracy, or criterion related validity, of the indices. One method to determine the accuracy of the indices would be to correlate them to an end of the year outcome measure. Likely, the strongest correlation between the indices and the end of the year outcome measures would be those that do not control for initial skills. Another method that could be utilized to determine the accuracy of the indices would be to investigate the percent of students making adequate progress in the indices and the percent of students moving across tiers of instruction in a three-tier model.

It would also be interesting to be able to identify the schools that maintained high rates of adequate progress for all their students over consecutive year periods. Mixing qualitative research and quantitative research with these schools could yield important information about curriculum, differentiated instruction, class size, intervention grouping size, and many other alterable variables that could be incorporated into other schools. Longitudinal research could also be conducted in order to determine student outcomes from these schools in later grades.

Overall, there were few schools that had high rates of adequate progress for the students scoring in the intensive, strategic, and barely-benchmark ranges. One avenue for future research would be to consider different patterns with the rates of adequate progress. For example, instead of just examining high rates of adequate progress over

semesters or years, middle rates and low rates of adequate progress in schools could also be examined. With the current study high rates of adequate progress was defined as the top 1/3 of schools or 80% of the students in each school that had adequate progress in the intensive and strategic ranges and the top 1/3 or 95% of the students in the barely benchmark range. A middle rate of adequate progress could be conceptualized as the middle 1/3 of schools or 50% of the students in each school that had adequate progress in the intensive and strategic ranges and the middle 1/3 of schools or 65% of the students in the barely benchmark range. Finally, a low rate of adequate progress could be conceptualized as the lower 1/3 of schools or 25% of the students in each school that had adequate progress in the intensive and strategic ranges and the lower 1/3 of schools or 35% of the students in the barely benchmark range. The three different ratings of adequate progress would be representative of the schools in the entire sample and may provide valuable information for schools. Including qualitative research about these schools would also be an interesting addition to the quantitative results.

In addition, the indices that are focused solely on intensive, strategic, or barely-benchmark students additional research can be conducted with those that make progress, but do not make adequate progress. Adequate progress decreased from the first semester to the second semester based on the results from the descriptive statistics for the indices that accounted for initial skills. In that second semester, it is likely that students are making progress, but not making adequate progress in their reading scores. Perhaps these students had made adequate progress in the first semester and to maintain that growth rate over the second semester was difficult with the instruction that the schools were

providing. The addition of a category below adequate progress (i.e., fair progress) may help to create more stable indices for the semester-to-semester time periods and still be reflective of students' growth.

If more sophisticated modeling techniques are utilized, linkages can be made from demographic characteristics (initial skills and free and reduced lunch) to the school-wide indices of adequate reading progress. The addition of these characteristics may account for potential differences between schools ability to support their students in making adequate reading progress. For example, if a school has a large percentage of students eligible for free and reduced lunch this may indicate that the school is located in a lower income area. That particular school may also have fewer resources available to dedicate to differentiated reading interventions and this may affect adequate reading progress in their students. Another factor that could be added to a model that may potentially impact the indices of adequate reading progress is instructional characteristics, such as the use of a scientifically based curriculum, teacher experience, or time spent on reading instruction during the school day. However, in order to include instructional characteristics independent data collection would have to occur instead of utilizing the existing data set from DDS since instructional characteristics are not inputted with DIBELS scores.

The final implication for future research would be to examine other variations of indices of school-wide adequate reading progress besides the six addressed with this study. It would be interesting to remove the students scoring in the well-above benchmark range for the outcome percent established index and the percent adequate progress index to order to create new indices to determine how the well-above

benchmark scores affected the stability of the current indices. It is likely that the stability of the new indices for outcome percent established and percent adequate progress with the well-above benchmark scores removed would mirror those that controlled for initial skills in the present study. In the future, the development of indices of evaluating adequate reading progress at the school level must have as much of a research base as the instruction that is being provided to students and the individual response to that instruction.

Conclusion

The present study provided initial evidence that the six proposed indices of school-wide adequate reading progress demonstrated varying degrees of stability, ranging from .10 to .90, depending on the index being examined. The indices that included all students within the school tended to exhibit higher levels of stability than those that controlled for initial skills level. However, the stability of the indices with all students within the school may have been confounded with the initial skill level that each school had. It also found that the year-to-year indices also tended to be more stable than the semester-to-semester indices. Since indices of school-wide reading progress hadn't been previously examined in the literature many directions of future research exist. As schools continue to work toward providing their students with differentiated instruction in order to support reading growth in their students, it becomes more imperative that indices are created to provide schools with a quick and efficient method of measuring of school-wide adequate reading progress. Continued pursuit of this line of research may be one method of providing that measurement. As research moves forward evaluating adequate progress

for reading, other academics, and behavior in addition to evaluating effective instruction it is important to remember to focus on features that are relevant and applicable in the applied setting of a school. Also, if RTI models for improving outcomes and evaluating eligibility for special education are utilized, there is a need for accurate and reliable indices of school-wide rates of adequate reading progress that inform the general effectiveness of 3 tiers of instruction. Indices of school-wide effectiveness will be required to have as much research support as the instruction, intervention, and assessment necessary for the three-tiered and RTI models.

REFERENCES

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Ashville, NC: North Carolina Teacher Academy.
- Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., Reschly, D. J., Schrag, J., & Tilly, W. D., III. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA : National Association of State Directors of Special Education, Inc.
- Buck, J. and Torgesen, J. (2003). *The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test*. Tallahassee, Fla.: Florida Center for Reading Research.
- Carlson, D. (2002) *Focusing State Educational Accountability Systems: Four Methods of Judging School Quality and Progress* (Dover, NH: National Center for the Improvement of Educational Assessment). Available online at: <http://www.nciea.org/publications/Dale020402.pdf>, accessed December 9 2008.
- Children's Educational Services, I. (1987). *Test of reading fluency*. Minneapolis, MN: Author.
- Daly, E., Martens, B., Barnett, D., Witt, J., and Olson, S. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review*, 36, 562-581.
- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Deno, S. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, 15 (3), 358-374.
- DIBELS Data System Description and Sample Reports (n.d.). Retrieved December 9, 2008, from <https://dibels.uoregon.edu/samples/index.php>.

- Duran, A. (2005). Factors to consider when evaluating school accountability results. *Journal of Law and Education, 34* (1), 73-100.
- Fuchs, L., & Fuchs, D. (1986). Curriculum-based assessment of progress toward long term and short term goals. *Journal of Special Education, 20*, 69-82.
- Fuchs, L. S. & Fuchs, D. (2006). A framework for building capacity for responsiveness to intervention. *School Psychology Review, 35* (4), 621-626.
- Fuchs, L. S., & Fuchs, D. (2006). Implementing responsiveness to intervention to identify learning disabilities. *Perspectives, 32* (1), 39-43.
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disability by responsiveness-to-instruction: Specifying measure and criteria. *Learning Disability Quarterly, 27*, 216-227.
- Fuchs, L., Fuchs, D., & Compton, D. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71* (1), 7-21.
- Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*, 429-438.
- Fuchs, L., Fuchs D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.
- Good, R., Simmons, D., & Kame'enui, E. (2001). The importance and decision making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5* (3), 257-288.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th ed.)*. Eugene, OR: Institute for the Development of Educational Achievement.
- Good, R., Wallin, J., Simmons, D., Kame'enui, E. J., & Kaminski, R. (2002) *System-wide percentile ranks for DIBELS benchmark assessment* (Technical Report). Eugene, OR: University of Oregon.
- Gravois, T., & Gickling, E. (2002). Best practices in curriculum-based assessment. In A Thomas, & J Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 885-898), Bethesda, MD: The National Association of School Psychologists.

- Gresham, F. M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review*, 33, 326-343.
- Gresham, F.M., VanDerHeyden, A., & Witt, J. (2005). Response to intervention in the identification of learning disabilities: Empirical support and future challenges. Unpublished manuscript. Available online at [http://www.joewitt.org/Downloads/Response to Intervention MS Gresham Vanderheyden Witt.pdf](http://www.joewitt.org/Downloads/Response%20to%20Intervention%20MS%20Gresham%20Vanderheyden%20Witt.pdf)
- Heller, K., Holtzman, W., & Messick, S. (Eds.). (1982). *Placement of children in special education: A strategy for equity*. Washington, DC: National Academy Press.
- Individuals with Disabilities Education Improvement Act (IDEA) of 2004*, 20 U.S.C. 1401 (c)(5)(F) (2004)
- Kaminski, R., & Good, R. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25 (2), 215-227.
- Kratochwill, T., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementation and sustaining multitier prevention models: Implications for response to intervention. *School Psychology Review*, 36, 618-631.
- Landis, J. and Koch, G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Linn, R. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40 (6), 699-711.
- Linn, R., Baker, E., & Betebenner, D. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31 (6), 3-16.
- Martson, D.B. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it? In M.R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: The Guilford Press.
- McIntosh, K., Chard, D., Boland, J., & Horner, R. (2006). Demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavior Interventions*, 8 (3), 146-154.

- Mellard, D. F., & Johnson, E. S. (2008). *RTI: A practitioner's guide to implementing response-to-intervention*. Thousand Oaks, CA: Corwin Press.
- Merrell, K. W., & Buchanan, R. (2006). Intervention selection in school-based practice: Using public health models to enhance systems capacity of schools. *School Psychology Review, 35*, 167-180.
- National Center for Education Statistics (2009). *The Nation's Report Card: Reading 2009 (NCES 2010-458)*. Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- No Child Left Behind Act of 2001*, 107th Congress 107-110 (2002).
- Reschly, D. J. (2005). Learning disabilities identification: Primary intervention, secondary intervention, and then what? *Journal of Learning Disabilities, 38*, 510-515.
- Shaw, D., & Shaw, R. (2002). *DIBELS Oral Reading Fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)* (Technical Report). Eugene, OR: University of Oregon.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research and Practice, 28*, 137-146.
- Viera, A. J. and Garrett, J. M. (2005) Understanding interobserver agreement: The kappa statistics. *Family Medicine, 37*(5), 360-363.
- Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency to performance on Arizona Instrument to Measure Standards (AIMS)* (Technical Report). Tempe, AZ.