

2008

# Qualitative analysis of teacher perceptions and use of the dynamic indicators of basic early literacy skills (DIBELS) within a district-wide Reading First program

Brian T. Gaunt  
*University of South Florida*

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

 Part of the [American Studies Commons](#)

## Scholar Commons Citation

Gaunt, Brian T., "Qualitative analysis of teacher perceptions and use of the dynamic indicators of basic early literacy skills (DIBELS) within a district-wide Reading First program" (2008). *Graduate Theses and Dissertations*.  
<http://scholarcommons.usf.edu/etd/258>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Qualitative Analysis of Teacher Perceptions and Use of the  
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

Within a District-Wide Reading First Program

by

Brian T. Gaunt

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

Co-Major Professor: Kelly A. Powell-Smith, Ph.D.  
Co-Major Professor: Michael J. Curtis, Ph.D.  
Robert M. Friedman, Ph.D.  
Roger A. Boothroyd, Ph.D.

Date of Approval:  
April 25, 2008

Keywords: data utilization, education, assessment, data-based decisions, evaluation

© Copyright 2008, Brian T. Gaunt

## Acknowledgements

In acknowledgement of all those who provided me with so much support, guidance, and encouragement throughout this process I wish to thank the following: Kelly Powell-Smith of Dynamic Measurement Group; Michael Curtis of USF School Psychology Program; Roger Boothroyd of the Louis De La Parte Florida Mental Health Institute; Robert Friedman of the Louis De La Parte Florida Mental Health Institute, Sharon Hodges of the Louis De La Parte Florida Mental Health Institute; Nancy Deane of Pinellas County Schools; Deanna Texel of Pinellas County Schools; my colleagues Bradley Beam and Janine Sansosti; my parents Susan and Thomas Gaunt; and of course my wife Maria and my daughters Viktoria and Natalia whose love and support were a foundation for all that I have accomplished.

## Table of Contents

Table of Contents	i
List of Tables	vi
List of Figures	vii
Abstract	viii
Chapter One – Introduction	1
Federal Response to National Reading Concerns	2
Reading First Program	3
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	4
Progress Monitoring and Reporting Network	7
Teacher Decision Making and Utilization of DIBELS data	12
Purpose of Present Study	13
Chapter Two – Research Review	16
History of Evaluation Research	16
Program Evaluation Research	17
Process Evaluation/Formative Evaluation	22
Rationale for Conducting a Process Evaluation	25
Process Evaluation in Education	29
Qualitative Research Methodology	35
Value and Nature of Qualitative Research Approaches	35
Purpose and Research Questions	39
Qualitative Research Methods	41
Researcher’s Biography and Role	41
Selection Methods	42
Data Collection and Analysis	45
Reliability and Validity	49
Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	51
Reliability and Validity of DIBELS	59
DIBELS Best Practice	68
Teacher Training and Data Utilization	86
Response to Intervention	88
Chapter Three – Method	90

Purpose	90
Research Questions	90
Research Design	91
Participants and Sites	92
Description of Sites	92
Description of Participants	92
Sampling Methods and Rationales Used for Selecting Participants and Sites	93
Selecting the School District	93
Selecting School Sites	94
Selecting Teachers, Reading Coaches, and Specialists	94
Sample Size	95
Teachers	95
Experts	97
Focus Groups	97
Data Collection Instruments	97
Semi-Structured Interview Guide – Teachers	97
Case Study – Teachers and Experts	98
Description of PMRN reports	99
Focus Group Guide	100
Instrument Validation Procedures	101
Data Collection and Analysis Procedures	102
Site Entry	102
Obtaining Consent to Participate	104
General Procedures for Obtaining Consent	104
Obtaining Consent from Reading First Supervisor	104
Obtaining Consent from Teacher Participants	105
Obtaining Consent from DIBELS Experts	105
Obtaining Consent from Reading Coaches	106
Obtaining Consent from Specialists	106
Researcher Biography	106
Background Experience	107
Researcher Beliefs and Expectations	107
Data Collection Procedures	109
General Procedures	109
Teacher Interviews	110
Teacher Case Study Review	111
Expert Opinion Interviews	113
Focus Groups	113
Data management and Storages	114
Data Analysis	115
General Data Analysis Procedures	115
Teacher Interview Concurrent Data Analyses	116
Teacher Interview Formal Data Analyses	117

Expert/Teacher Case Reviews	124
Focus Group Analyses	126
Member Checks and Peer Reviews	127
Chapter Four – Results	129
General Introduction	129
Description of Participants and Data Obtained	129
Teachers’ Perceptions and Understandings About DIBELS and PMRN	130
Climate and Culture of Schools	130
Support/Resources for Teachers’ Use of DIBELS	136
Knowledge of DIBELS	141
Correspondence With Other Assessments	146
DIBELS and Progress Monitoring	154
Using Assessment Data	155
Comparison of Teachers and Experts on the Use of PMRN Reports - Case Study Review	161
Expert Review of Kindergarten Case Study	162
Kindergarten Teachers’ Review of Kindergarten Case Study	164
Expert Review of First Grade Case Study	165
First Grade Teachers’ Review of First Grade Case Study	166
Expert Opinions on the Use of DIBELS at <i>Reading             First Schools</i>	168
Attitudes and Perceptions Among Persons Other Than Teachers	171
Climate and Culture of Schools	171
Supports/Resources Available to Support Use of DIBELS	174
Teachers’ Knowledge of DIBELS	177
Collecting and Using Assessment Data	181
Analysis of Hypotheses and Researcher Expectations	188
Analysis of Unanticipated Topics	191
Analysis of Qualitative Research Process – Researcher’s Reflections	195
General Introduction	195
Research Proposal and Preparations for Conducting Study	195
Accessing the Research Site and Participants	197
Data Collection and Data Analysis	202
Chapter Five – Discussion	214
Discussion of Primary Results	216

Context/Culture/Climate	216
Resources/Supports/Infrastructures	225
Limitations	236
Contributions of the Present Study	240
Final Summary	241
Create a Culture of Change and Shared Vision	241
Create Capacity and Infrastructure to Support Use of DIBELS	243
Provide Additional and Ongoing Training on the use of DIBELS/Formative Assessments	245
Provide Educators with Ongoing Supports for Data Analysis and Utilization	246
References	249
Appendix A – Figures	261
Figure 1 – Example of Class Status Report	262
Figure 2 – Example of Student Grade Summary Report	263
Figure 3 – Example of Reading Progress Monitoring Student Cumulative Report	264
Figure 4 – Instructionally Interpretable Zones of Performance	265
Appendix B – Tables	266
Table 3 – Number of Schools Sampled Across the Three Groups of Participants	267
Table 4 – Probability Sampling Table Developed by DePaulo (2000)	268
Table 5 – Microsoft Excel Format Used for Organizing, Coding, and Sorting Transcript Data	269
Table 6 – Comparison of Themes Identified Across Multiple Sources to Organize Findings for Presentation	270
Table 7 – School District Assessment Schedule for Kindergarten and First Grade	271
Appendix C	272
Appendix C1 – Teacher Interview Guide	273
Appendix D	275
Appendix D1 – Case Study Questions	276
Appendix D2 – Case Studies	277
Appendix E	283
Appendix E1 – Focus Group Guide	284

Appendix F	285
Appendix F1 – Field Notes Teacher Interviews Format	286
Appendix G	288
Appendix G1 – Teacher Interview Code Draft 1	289
Appendix G2 – Teacher Interview Codes Draft 2	290
Appendix G3 – Teacher Interview Codes Draft 3	295
Appendix G4 – Teacher Interview Codes Draft 4	300
Appendix H	305
Appendix H1 – Preliminary Results for Member Checks and Peer Review	306

## List of Tables

Table 1	DIBELS benchmark screening for each grade level	6
Table 2	General PMRN reports for use	10
Table 3	Number of schools sampled across the three groups of participants	272
Table 4	Probability Sampling Table for Selection of Participants	273
Table 5	Microsoft Excel format used for organizing, coding, and sorting transcript data	274
Table 6	Comparison of themes across multiple sources to organize findings for presentation	275
Table 7	School district assessment schedule for kindergarten and first grade	276

## List of Figures

Figure 1	Example of a Class Status Report	267
Figure 2	Example of a Student Grade Summary Report	268
Figure 3	Example of a Reading Progress Monitoring Student Cumulative Report	269
Figure 4	Graph developed by Good et al. (2000) for use in a Benchmark Linkage Report	270

A Qualitative Analysis of Teacher Perceptions and Use of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Within a District-Wide Reading First Program

Brian T. Gaunt

ABSTRACT

The aim of the *Reading First* grant program was to (a) increase quality and consistency of instruction in K-3 classrooms; (b) conduct timely and valid assessments of student reading growth in order to identify students experiencing reading difficulties; and (c) provide high quality, intensive interventions to help struggling readers catch up with their peers (Torgesen, 2002). In the State of Florida, school districts must incorporate the use of an assessment tool called the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) to qualify for *Reading First* grant funding. Though DIBELS has been found to be a valid and reliable assessment for screening, monitoring, and evaluating student outcomes in early literacy skills, very little discussion or research has been conducted concerning teacher use and attitudes about DIBELS within a *Reading First* program. The present study involved a qualitative analysis of teachers' perceptions and use of the DIBELS within a Reading First context. Fourteen teachers (seven kindergarten and seven first grade teachers), Reading Coaches, non-teaching Specialists, and DIBELS experts participated in the present study. Results were aggregated for comparisons across multiple data sources. Results suggest teacher's perceptions may not be easily classified on a simple dichotomous range; rather their reported benefits and concerns on the use of

the DIBELS were found to be varied and highly situational. Results were further interpreted in the context of research literature on data utilization and analysis in schools.

## CHAPTER ONE - INTRODUCTION

Reading is perhaps the most important academic skill for children to acquire. Indeed it is the cornerstone upon which many other academic and cognitive skills develop for school-aged children. Given its central importance in the socialization and education of children, it is troubling to discover how many students are affected by illiteracy. Moats (1999) found that approximately 20% of elementary students nationwide have significant problems learning to read. At least 20% of elementary students do not read fluently enough to enjoy or engage in independent reading. The rate of reading failure for African-American, Hispanic, limited-English speakers and children living in poverty ranges from 60% to 70% (Moats, 1999). One-third of poor readers nationwide are from college-educated families. Twenty-five percent of adults in this country lack the basic literacy skills required in a typical job.

The statistics concerning illiteracy in the United States are so pervasive that the National Institutes of Health regarded reading development and reading difficulty as a major public health concern (Moats, 1999). However, the convergent literature has convinced many leaders in reading research to conclude that reading failure is unnecessary and can be prevented and ameliorated. It is also important to note that experts suggested reading instruction is a significant challenge for teachers (National Reading Panel, 2000). Unlike spoken-language acquisition, reading is not a natural process. Rather, many researchers argue reading instruction must involve explicit, direct,

and systematic instruction (Foorman, Francis, Fletcher, Paras, & Schatschneider, 1998; Stanovich & Stanovich, 1995), and teachers must be provided specialized training and preparation to teach reading (National Reading Panel, 2000).

### *Federal Response to National Reading Concerns*

Though research on early intervention in literacy is several decades old, a new wave of public discussion and debate at the federal level of education and policy making has been occurring. Much of the research in early literacy intervention argues reading development begins early in a child's life and therefore requires explicit and systematic focus in the kindergarten through third grade years (e.g., Adams, 1990) for successful reading development. The perceived convergence of research (e.g., National Reading Panel, 2000) on early identification and prevention of reading difficulties may have led to some of the increased emphasis in current policies enacted which place greater attention on accountability and focus on the use of "scientifically-based" instruction in reading.

One such policy enacted by the federal government was the No Child Left Behind Act (NCLB), which was signed into law by President Bush on January 8, 2002. The act was based on the following four principles: (1) stronger accountability for results, (2) more freedom for states and communities, (3) encouraging proven educational methods, and (4) more choices for parents. At the core of this legislation was a greater emphasis on scientifically-based teaching of early literacy skills in kindergarten through third grade.

### *Reading First Program*

The *Reading First* grant program was a core component of the NCLB Act. This grant program, which provided the nation's schools with approximately \$900 million in 2003, was intended to promote the use of scientifically-based research to develop high-quality reading instruction for grades K-3. The three main funding categories available within a *Reading First* grant are professional development for teachers, purchase and implementation of assessments, and purchase of materials including software and books. The program was "...designed to select, implement, and provide professional development for teachers using scientifically-based reading programs, and to ensure accountability through ongoing valid and reliable screening, diagnostic, and classroom-based assessment ([www.ed.gov/programs/readingfirst/index.html](http://www.ed.gov/programs/readingfirst/index.html), 2003)." More specifically, the aim of the *Reading First* program was to (a) increase quality and consistency of instruction in K-3 classrooms; (b) conduct timely and valid assessments of student reading growth in order to identify students experiencing reading difficulties; and (c) provide high quality, intensive interventions to help struggling readers catch up with their peers (Torgesen, 2002).

On September 7, 2001, Florida Governor Jeb Bush signed legislation (Executive Order #2001-260) to create the *Just Read, Florida!* initiative which was designed to coordinate several programs and grant funded initiatives already in place, while also providing a comprehensive plan to increase literacy rates in Florida's public schools. This legislation created a leadership "triangle" to support and monitor the implementation efforts of this legislation ([www.justreadflorida.com/docs/guidance.pdf](http://www.justreadflorida.com/docs/guidance.pdf)) which consisted

of the Florida Department of Education (FLDOE), the Florida Center of Reading Research (FCRR) and the Florida Literacy and Reading Excellence (FlaRE) center.

In addition to providing leadership direction for school districts by developing and evaluating research-based reading curriculums and instructional practices, FCRR provided technical assistance to school districts that have been awarded *Reading First* grants. The technical assistance provided by FCRR involved, but was not limited to, consultation about appropriate assessment tools for inclusion in a school district's *Reading First* plan, training on the use of progress monitoring and outcome assessments, and management/consultation for data collection and analysis of assessment data. Additionally, FCRR coordinated with the FlaRE center to identify professional development needs such as appropriate instructional materials and resources, and professional development materials and practices that are supported by scientific knowledge about reading and professional development ([www.justreadflorida.com/docs/guidance.pdf](http://www.justreadflorida.com/docs/guidance.pdf)).

#### *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*

A key implementation component of the *Reading First* program plan in Florida was the use of DIBELS (Torgesen, 2002). In general, a school district in Florida must submit a plan which describes coordination of four types of assessments to be eligible to receive grant funding. These four types of assessments are screening, diagnostic, progress monitoring, and outcome assessments. The *Reading First* program in Florida required each school district to commit to using DIBELS. Though school districts may supplement with other assessment data, DIBELS purported to offer a cost effective, time

efficient approach that offered educators a reliable and valid means to (1) screen students to determine risk for reading failure, (2) progress monitor student achievement, and (3) assess outcomes in early literacy skills (Elliott, Lee & Tollefson, 2001; Kaminski & Good, 1996).

The DIBELS provide very brief, individually administered tests of critical early pre-reading and reading skills (Kaminski, Cummings, Powell-Smith, & Good, 2008). In Florida, this assessment was intended to be administered at least three times a year as a screening tool, though schools may administer DIBELS more frequently to monitor student progress more closely. More frequent progress monitoring may also be especially helpful for children who are assigned to receive more intensive reading instruction because of the importance of monitoring closely the effectiveness of interventions for students who are at-risk of reading failure (Kaminski et al., 2008).

In general, DIBELS, as utilized in the State of Florida, involved five subtests used at specific grade levels. The five subtests are letter naming fluency (LNF), initial sounds fluency (ISF), phoneme segmentation fluency (PSF), nonsense word fluency (NWF), and oral reading fluency (ORF). Table 1 depicts the FCRR recommended schedule for grades K-3 in Florida. This table shows the recommended school days in which DIBELS would be utilized as a screening instrument and the subtests used for each grade level.

Table 1  
DIBELS Benchmark Screening Schedule for Each Grade Level.

	Days 20-30	Days 65-75	Days 110-120	Days 155-169
Kindergarten	LNF ISF	LNF ISF	LNF ISF PSF NWF	LNF PSF NWF
First Grade	LNF PSF NWF ORF	PSF NWF ORF	PSF NWF ORF	PSF NWF ORF
Second Grade	NWF ORF	NWF ORF	NWF ORF	NWF ORF
Third Grade	ORF	ORF	ORF	ORF

The term “fluency” in each of the subtests indicates a measure of speed plus accuracy about a student’s progress. With exception to the Initial Sound Fluency subtest, all subtests in the DIBELS involve a one-minute timed measure. Letter Naming Fluency (LNF) is a measure of how fast and accurate children can say the names of letters printed on a page. Students are shown upper and lower case letters that are arranged in random order, and they are asked to name as many letters as they can in one minute.

Initial Sound Fluency (ISF) is a measure of early phonemic awareness. Children are presented with four pictures at a time and asked to point to the picture which corresponds with the initial sound provided orally by the examiner (for three of the pictures) and say the initial sound of the fourth picture. This subtest measures the cumulative latency, in seconds, of student responses to task questions asked by the examiner.

Phoneme Segmentation Fluency (PSF) is a slightly more advanced measure of phonemic awareness. It tests children’s ability to pronounce and segment the individual phonemes in words that have three and four phonemes (e.g., cat or rest). Children are orally provided with a word and asked to say all the sounds they hear in the word.

Nonsense Word Fluency (NWF) measures children’s knowledge and skill in applying the alphabetic principle. Children can earn credit either by giving the individual sounds represented by the letters in simple non-words (e.g., /l/-/u/-/t/) or by blending the sounds together and pronouncing the non-word as a whole word (e.g., lut).

Oral Reading Fluency (ORF) (also called CBM-Reading—Good, Simmons, & Kame’enui, 2001) is a measure of children’s ability to read grade level text aloud fluently and accurately. Children receive a score based on the number of words in a grade-level passage they can read accurately in one minute.

#### *Progress Monitoring and Reporting Network*

In Florida, schools utilized DIBELS at least three times during the school year in order to identify and monitor the students’ progress in reading development. The DIBELS data were submitted into a web-based data management program known as the Progress Monitoring and Reporting Network (PMRN) which was developed and is managed by FCRR. This reporting system was designed to provide school districts with timely and reliable information about student performance on screening, progress monitoring, and outcome assessments in order for teachers to effectively plan classroom instruction and to assist schools in the evaluation of their core curricula and instructional practices. The PMRN system, combined with the assistance of reading coaches at each

school—an intended component of the *Reading First* model in Florida—offered teachers a potentially valuable approach to immediately identifying students who are struggling, identifying areas of instructional need when designing reading interventions, and evaluating classroom instructional and curriculum variables.

All K-3 student data in all *Reading First* schools were required to be included in this system of data collection; though fourth and fifth grade data may also be entered and utilized using the PMRN. Specifically, the purpose of the PMRN system , as indicated on the Florida Center for Reading Research website, was to “efficiently and accurately accomplish three tasks: (1) allow the data from required tests to be entered quickly and easily; (2) store the data in a safe and secure location; and (3) provide timely and helpful reports to educators ([www.fcrr.org](http://www.fcrr.org), 2004). And, according to the *Reading Coach’s Guide* (August, 2003) provided by FCRR, the *Reading First* program in Florida requires that information from assessments guide reading instruction and this information should inform teachers about the following:

1. Risk level of students
2. Specific weaknesses in reading skills that one or more students exhibit
3. How students should be grouped
4. The intensity of instruction required for certain students through small groupings or more time on skill-building activities
5. Which skills should be emphasized for particular students
6. How much change is occurring in student skills over time following interventions

7. The professional development needs of the teacher in the area of reading instruction

There are four basic steps to using the PMRN system: (1) assessment teams (made up of professionals at the school and trained by FCRR regional coaches) collect the DIBELS data, (2) designated data entry personnel input the student data into the PMRN system, (3) the PMRN system quickly aggregates the data and allows reports to be generated, and (4) teachers and other educators access the PMRN reports by accessing the web-site utilizing a password and username unique to the teacher. Because the web-site was password protected, teachers had access to aggregate data for their classrooms as well as to individual student reports for students in their classroom. Similarly, administrators generally had access to aggregated data for their school, aggregated data of each classroom, and aggregated data by grade level. District leaders generally had access to aggregated data of the district and aggregated data for each school and grade levels for the district.

A variety of reports were available to teachers using the PMRN system. In general, a student's data would have been displayed in any of three levels of reports: individual, class, and school. Table 2 lists the general types of information that can be found in PMRN reports. Throughout the reports, a color code was used to provide a visual representation of students' instructional needs and risk-level. Red indicated immediate intensive intervention was needed and the student was at high-risk of not achieving grade-level reading skills. Yellow indicated additional instruction was needed to improve targeted skill areas and the student was identified as having a moderate risk

status. Green represented the student was likely to achieve grade level reading skills with use of the core curricula and instruction and was subsequently identified as having a low-risk status. Blue represented above average performance on a particular measure or identified a student as having above grade level reading skills.

Table 2  
General PMRN Reports for Use

Report	Description
Progress Reports:	Shows gains students are making in a specific subtest area and how the performance has improved since the last assessment.
Summary Reports:	Show the percentage of students who may need extra instruction compared to those who are making adequate progress.
Historical Reports:	Compare the progress of the current year's class to that of previous year's classes.
Comparison Reports:	Compare the progress of a class or school to others in Florida serving similar children. These are only available to schools at which all students are being Progress Monitored.
Cumulative Reports:	Shows the results of all reading assessments recorded in the PMRN across all years and Year End Outcome Test scores, including scores from other schools.

Appendix A contains the three types of specific reports that were used in the present study. The Class Status Report (Figure A1) was a common progress report used by teachers which showed how well students in the class performed on a specific assessment cycle using DIBELS. The list of students can be organized in either alphabetical order, level of instruction recommended (e.g., intensive, strategic, or initial), or by subtest score. The recommended instructional level was a feature of the PMRN

that was generated for the teacher to direct their use of DIBELS data for developing broad instructional interventions. The colors used by the PMRN system were intended to convey a level of instructional support necessary for helping a student achieve end-of-year outcomes. Red indicated that intensive remediation was necessary (i.e., High-Risk; HR), yellow indicated that moderate remediation supports were needed (i.e., Moderate-Risk; MR), green indicated that general classroom curriculum was sufficient for continued progress (i.e., Low Risk; LR – core curriculum considered sufficient), and blue indicated that a student was performing above average (i.e., Above Average; AA).

A commonly used summary report was the Student Grade Summary Report (Figure A2) which showed the progress of an individual student for a single assessment cycle for all DIBELS measures used for that assessment cycle. This report allowed the classroom teacher to analyze a specific student's status in comparison to the normative distribution of the classroom for the same assessment cycle. By using a box plot format, the upper and lower limits of the box indicated the 75<sup>th</sup> and 25<sup>th</sup> percentiles, respectively, while the middle horizontal line within the box indicated the median population score. Individual data points above or below the whiskers indicated outliers whose scores fell within the upper or lower 5<sup>th</sup> percentiles. The individual student's score that was being analyzed was represented by a colored flag with the corresponding letters indicating that student's overall status (e.g., HR = High Risk).

The Reading Progress Monitoring Student Cumulative Report (Figure A3) allowed a teacher to review a student's progress across all cycles given throughout the year in addition to having student scores for end-of-year outcomes on other assessments

such as the Peabody Picture Vocabulary Test (PPVT), the Stanford-10 Achievement Test, and the Florida Comprehensive Assessment Test (FCAT). Though a color-coding system was not utilized in this report, the risk indicator labels (e.g., HR) were available. One unique feature of this report was the use of a Recommended Instructional Level (RIL) rating which depicted a student's overall instructional level for a given assessment cycle. Analysis of this indicator across all assessment cycles allowed a teacher to identify the overall trend in a student's performance throughout the school year.

#### *Teacher Decision-Making and Utilization of DIBELS Data*

The issue of decision making and use of assessment information by teachers is a critical issue that has been addressed in research concerning the use of Curriculum-Based Measurement-Reading (Wesson, Skiba, Sevcik, King, Tindal, Mirkin, Deno, 1983) which is closely related to the DIBELS measures (Good et al., 2001). Though the technical adequacy on DIBELS have been researched and found to have acceptable reliability and validity for measuring students' early literacy skills (e.g., Good, Kaminski, Simmons, & Kame'enui, 2001), little is known about how teachers are utilizing the assessment data to guide instructional decision-making within a *Reading First* context. Knowing this information is important because the degree to which a classroom teacher uses an assessment tool as it was designed to be used and interpreted can have a significant impact on outcomes (Wesson et al., 1983).

For example, Wesson et al. (1983) trained three groups of teachers in the use of Curriculum-Based Measurement-Reading; each at different levels of technical adequacy (high, medium, and low). The study examined the effects of the level of training as an

independent variable on student's reading achievement as measured by student performance on three reading passages using CBM-Reading measures. Their results showed students of those teachers who more accurately and consistently applied CBM made greater achievement gains. The key variable that was observed by the researchers was the degree to which teachers made instructional decisions based upon technically adequate data and consistent rules regarding how to use this data.

Other studies have found much diversity among teachers in regard to mastery of CBM measurement and data analysis techniques (King, Deno, Mirkin, & Wesson, 1983; Skiba, Wesson, & Deno, 1982), suggesting some teachers either need extensive training or ongoing support in order to use the measurement data in a valid and meaningful way. Although reading coaches—individuals assigned to a single school for providing ongoing consultation and training for teachers—are a core component of the *Reading First* program in Florida, little was known about how these coaches impacted teacher use of DIBELS data, or how teachers perceived the value of the reading coaches in their school.

#### *Purpose of Present Study*

DIBELS offer a preventive approach to handling reading failure and are designed to be used within an Outcomes Driven Model (Kaminski et al., 2008) for reducing potential reading difficulties while supporting all children to achieve adequate reading outcomes by the end of the third grade (Good, et al., 2001). Researchers have shown that when used appropriately DIBELS can make a positive difference in a classroom where struggling readers may be found (Smith, Baker, & Oudeans, 2001).

When DIBELS are used in conjunction with teacher knowledge of reading instruction, as well as specific feedback about implementation during ongoing professional development, significant improvements in student reading achievement may be obtained (Smith, et al., 2001). As Baker, Smith, Kame'enui, McDonnell & Gallop (1999) stated, it is the “way the curriculum was implemented (and not the specific curriculum they used) [that] made the difference between successful and problematic learning for many [students].” DIBELS is a set of indicators that may be used to assess students' early literacy skills regardless of the curriculum program used. Thus, teachers are a fundamental and necessary component for successful use of DIBELS towards a goal of all children reading at grade level.

The primary purpose of the present investigation was to conduct a qualitative analysis of teacher's perceptions and use of DIBELS within a *Reading First* program. Several factors made such an investigation necessary. First, the research literature indicated that a teacher's appropriate use and understanding of assessment data was important for supporting and improving student reading outcomes. To date, no studies were found by this examiner which evaluated teacher's perceptions and understanding about using DIBELS within a *Reading First* context.

Second, evaluating process variables such as program participant beliefs, knowledge, and perceptions of a given program can supplement and assist in the evaluation of a program's outcomes. Program evaluation studies typically involve, at least, an evaluation of the program's outcomes in terms of value and effectiveness (Patton, 1990). However, several unanswered questions can still remain after an

evaluation of a program's outcomes. Though a complete process evaluation of the entire *Reading First* program was beyond the scope of the present study, an evaluation of a central component of the *Reading First* program was warranted (i.e., teacher's perceptions and use of DIBELS).

Third, though FCRR can monitor the frequency of teacher use of the PMRN system (i.e., frequency of log-in attempts), there did not appear to be any formal collection of information regarding teacher perceptions and value ascribed to using the PMRN. For example, what reports did teachers find most useful? Do teachers need further support or training regarding their use and understanding of PMRN reports? Finally, a qualitative approach to investigating teacher perceptions and use of DIBELS was best suited for answering process related questions by which specific events or activities occur and the meanings individuals place on those events or activities (Patton, 1990). Given these issues, the following research questions were proposed for investigation:

1. What are teachers' perceptions and understandings about DIBELS and the PMRN?
2. How do teachers' understandings and use of DIBELS data, as presented in the PMRN reports, compare to *Reading First* experts who are provided with the same information?
3. What attitudes and perceptions exist among persons other than teachers who participate in the collection, input, and analysis of DIBELS data throughout the school year?

## CHAPTER TWO – Research Review

The following section is a review of relevant research and discussion articles covering the essential elements and procedures for utilizing a qualitative approach to process evaluation for the improvement of the *Reading First* program. It is important to note in the beginning that published material on process evaluation involving prevention programs was limited. After a summary of process evaluation approaches, qualitative analysis procedures are reviewed. Next, a review of the literature concerning the use of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is provided, followed by a review of research investigating teacher use of assessment data, so that the reader has a context for applying a qualitative approach to process evaluation in the present study.

### *History of Evaluation Research*

Rossi and Freeman (1993) provided a sufficient review of the historical events leading to the modern day field of evaluation research. Evaluation research emerged from the general acceptance of the scientific method around the 17<sup>th</sup> century. Though its beginnings can be traced to this time period, evaluation research was considered a relatively modern development. The university setting provided for the earliest work on program evaluations and eventually led to an increased popularity in the social sciences. The fields of education and public health were among the first to use systematic evaluation efforts. Topics such as literacy rates, morbidity rates, and mortality rates in society were researched in earlier years. A sizable increase in the use of program

evaluations of community action programs was observed in the 1930's. After World War II, and the inception of many federally funded social programs, an accelerated pace in the use of program evaluation occurred to meet the needs of urban development and human need associated with urban living; similar rates in program evaluation studies also occurred in order to meet the demand for measurable results of large-scale programs.

By the 1950s large-scale evaluations were commonplace in society. During the 1960s, papers and books on evaluation research grew substantially in number. In the 1970's professional journals and conference organizations developed rapidly in response to the developments in program evaluation research. Much of the advances in evaluation research were due to advancements in statistics and related research methods. Today the role of evaluation research remains substantial. Regardless of political partisanship and leadership the role of evaluation research does not seem to change.

### *Program Evaluation Research*

Some researchers define evaluation as any activity directed at collecting, analyzing, and interpreting information on the need for, implementation of, and effectiveness and efficiency of intervention efforts to better the lot of humankind (Rossi and Freeman, 1993). Rossi and Freeman defined evaluation research as, "the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs" (p. 5). The purpose of evaluation according to the authors depended on the types of questions asked about a particular program. In general, evaluations are undertaken to (1) judge worth, (2) estimate usefulness to improve, (3) assess utility of new programs, (4) increase

effectiveness, and (5) satisfy accountability requirements and sponsors. Patton (1990) defined evaluation more broadly as any effort to increase human effectiveness through systematic data-based inquiry. He generally conceptualized evaluation efforts as actions that inform stakeholders and enhance decision-making to solve human and societal problems.

Whereas researchers such as Cook and Campbell (1979) argued for a purely scientific approach to evaluation research, others advocated for a more pragmatic approach similar to that offered by Cronbach (1982) (Patton, 1990; Rossi & Freeman, 1993). Evaluation research is different from scientific studies in purpose and intent (Cronbach, 1982; Patton, 1990; 2002; Rossi & Freeman, 1993). The political nature and context in which many evaluation studies evolve from and operate is one distinctive feature of evaluation research. Evaluation efforts should be idiosyncratic to meet the needs of program sponsors and stakeholders. Scientific studies strive to meet standards set by the investigators' peers and the overall statistical and theoretical rigor of standards imposed by research journals. Evaluations should be designed and implemented in ways that recognize the policy and program interests of the sponsors and stakeholders. Evaluations should yield useful information for decision makers given the political circumstances, program constraints, and available resources. "Resource limitations and more realistic expectations for social programs only increase the need for evaluation efforts as societies attempt to cope with their human and social problems" (Rossi & Freeman, 1993; p. 7).

The intent of providing evaluation information that is useful for stakeholders and policymakers is consistent with Chen and Rossi's (1983) observation of the difference between controlled scientific and socially motivated evaluations. Their observations suggested controlled experiments were an attractive and seductive approach to evaluating a program in that such studies allow for an estimation of net effects through randomized experiments provided that the goals and objectives of a program were stated in objective and measurable terms. However, defining a program's goals and objectives in objective and measurable terms is not always a salient option. That is why Chen and Rossi placed great importance on going beyond identifying the goals of a program to also identify the goals articulated by policy makers and other stakeholders who have legitimate interests in some aspects of the program. This distinction was important because outcome variables identified through identification of program goals tended to be narrower than the connotative intentions of program designers and/or policymakers that may not have been included in the original program design.

Legislation in the United States deliberately fosters diversity in the process of implementation, and therefore local conditions may require extensive adaptations in the program (Chen & Rossi, 1983). Where such adaptations are made, it is important to evaluate the intended and stated goals of the policy makers in comparison to the intended and unintended goals of those local personnel who would implement such a program. That is why Rossi and Freeman (1993) stated, "...evaluators must be responsive to the context in which they are working (p. 27)." The aim of an evaluator should be the

provision of the most reliable findings possible given the political and ethical constraints, and limitations imposed by time, money, and human resources.

With regards to conducting evaluation research, particularly where programs are being evaluated to determine level of success achieved, Chen and Rossi (1983) suggested categorizing program goals in one of three categories: (1) Policy-directed/Plausible Goals, (2) Policy-directed/Implausible Goals, and (3) Theory-driven/Not Specified by Policy Goals (Chen & Rossi, 1983). The overall purpose of making such distinctions is to not only confirm that intended effects are occurring, but also to identify effects that were not intended by their designers. Such unintended effects may be either desirable or undesirable and may offset those intended. An evaluator should take into account inferred effects as well as those directly intended. The first category identifies those effects the program designers intend to occur when the program is implemented as designed; where the goals and specifications for program implementation and operation are clearly defined and specified. The second category identifies program goals or effects that are intended by program designers, but likely implausible due to low specificity of implementation requirements and program operation. The third category refers to effects not specified through policy or program designers, but is nonetheless plausible where local adaptations are likely to occur.

A combined qualitative and quantitative approach to conducting evaluation studies may be utilized, and in some cases is best practice (Rossi & Freeman, 1993). These researchers argue the use of multiple sources increases the range of data collected and promotes data and methodological triangulation while increasing the validity of

conclusions. Specifically, the authors suggest using direct observations by the evaluator, written records, data from delivery agents of the intervention and data from intervention trial participants. Supportive of Chen and Rossi's (1983) emphasis on the evaluation of unspecified-plausible goals at the local implementation level, Rossi and Freeman suggest a mixed methodological approach to evaluation combined with data collection from multiple sources provides a means for detecting such goals and unintended program features.

Having considered the purpose of evaluations and the role of the evaluator working within a social and political context, evaluators must also take important steps to ensuring that the results of their evaluation are useful for policy makers of the program being evaluated and stakeholders who may have different questions about and expectations for a given program (Chen & Rossi, 1983; Rossi & Freeman, 1993). Thus, merely providing information does not ensure its use. Dissemination and utility of information may be conceptualized as more of an art form. It is important to understand the organizational arrangements under which a social program is operating. Evaluators should consider developing a strategic plan for providing feedback to policymakers and stakeholders if the information is to be found valuable for use. In the development of such plans, it may be necessary to establish a dissemination plan early in the evaluation process and revise the plan when new information yields specific requirements for ensuring stakeholders and policymakers use the evaluation information obtained. As such, evaluation may be viewed as an evolving process rather than a static series of steps to follow.

### *Process Evaluation/Formative Evaluation*

Some researchers suggest that process evaluations have traditionally not been given the attention they fully deserve in evaluation research (Chen & Rossi, 1983). However, this trend appears to be changing as more and more process evaluations are being conducted in a wide variety of social and political areas. Researchers typically characterize process evaluation as involving the monitoring of implementation and accountability practices of a program (Rossi & Freeman, 1993), examining the relationships among program components to identify critical intervention elements for future practitioners (Chen & Rossi, 1983), adherence to program implementation designs and specifications (Elliott & Mihalic, 2004), identifying a reasonable balance between program implementation fidelity and local adaptation of program implementations (CSAP, 2001), and identifying means for improving a program (Rossi & Freeman, 1993). Some have conceptualized process evaluation as a means to collect feedback about program implementation, participant responses to the program, changes to the site in response to implementation, and information about personnel competency (Israel, Cummings, & Dignan, 1995). Helitzer, Yoon, Wallerstein, Dow y Garcia-Velarde, (2000), argued process evaluation can provide critical documentation necessary for sustaining and replicating successful community-based trials. Process evaluation may be used as a key to understanding the internal dynamics of an intervention trial and to monitor the quality of a program.

Rossi and Freeman (1993) suggest three questions when undertaking a process evaluation. The first question concerns the extent to which a program is reaching the

appropriate target population. The second question concerns the delivery of services in terms of consistency with program design specifications. The third question asks what resources are being or have been expended in the conduct of the program. The second of Rossi and Freeman's questions is synonymous with the concept of fidelity. It has been argued that a valid evaluation study rests upon the consistency to which a program is implemented as designed and specified for use (Elliott & Mihalic, 2004). However, according to Chen and Rossi (1983) implementation research, or process evaluations, have focused too much at worrying about fidelity and not enough about understanding the process of implementation.

Fidelity is an important concept that often arises when the topic of process evaluation is discussed. A debate exists in the research literature about fidelity and local adaptation (Blakely, Mayer, Gottschalk, Schmitt, Davidson, Roitman, & Emshoff, 1987; Castro, Barrera, & Martinez, 2004; Elliott & Mihalic, 2004; Kovaleski, Gickling, Morrow, & Swank, 1999). According to Elliott and Mihalic (2004), there has been a long standing debate between the need for implementing programs as they were designed and the need to adjust program aspects to fit the local conditions. Some researchers have argued evaluations should seek a balance between fidelity and local adaptation of a program's implementation (CSAP, 2001), while others have questioned the assumption that local adaptation is inevitable (Elliott & Mihalic, 2004).

The issue of fidelity as it applies to evaluating staff persons who participate in the implementation and use of a program typically involves a combination of three factors: setting conditions, adherence, and competence (Fixsen, Blasé, Naoom, & Friedman,

2005). With respect to the present study, these factors are relevant for consideration. According to Fixsen et al., setting involves structural aspects of a program that must be in place for program success. Adherence is the extent to which staff persons use the program in a manner intended by the program designers. Competence refers to the skills one has in utilizing components of the program as designed and intended for use.

Elliott and Mihalic (2004) suggest an appropriate amount of support and encouragement from directors, or administrators, is essential in motivating staff to adhere to program features as designed. The authors argue that because the research literature has demonstrated the importance of fidelity and its significant relationship to effectiveness in outcomes any bargaining away from fidelity will lead to a decrease in program effectiveness. Elliott and Mihalic indicated their observations suggest local adaptation that occurred incidentally in their implementation efforts did so at the direct staff or teacher level and not at the administrator level. These adaptations were typically made without input from designers or implementation consultants. They suggest that where such adaptations are occurring it is important to have a process for identifying them and determining the degree to which they may affect outcomes.

Given the relevant research available about process evaluations, it is important to recognize that just as it is important to identify the program's goals when evaluating outcomes of a program, it is important to identify the purpose of an evaluation in order to determine the relevant dependent variables to be measured during a process evaluation. If stakeholders and policymakers explicitly do not intend or desire local adaptations to occur, then utilizing a process evaluation for measuring fidelity of program

implementation and use would be best practice. However, if adaptation and flexibility are built into the design of a program, then the issue of fidelity may not be as important as identifying the factors that lead to successful outcomes for the program as a whole, as well as factors that improve staff adherence to certain components of the program and competence in using the program.

### *Rationale for Conducting a Process Evaluation*

Regardless of what the intent is of conducting a process evaluation, there are several potential benefits in conducting a process evaluation. Evaluation of a program's outcomes alone often provides narrow and sometimes distorted understandings of the program (Chen & Rossi, 1983). Chen and Rossi provided a theory-driven approach to conducting evaluations by conceptualizing evaluation research as the testing of hypotheses for why a particular program or service is failing or succeeding. The authors argued that a program's failure might be due to poor conceptual foundations, low dosage levels of intervention, or because of poor implementation (e.g., fidelity).

When outcome evaluations reveal little or no impact it is still possible to estimate if a program's treatments are efficacious by examining process evaluation data (Rossi & Freeman, 1993). Process evaluations can provide information about efficiency of operations. Process evaluations also provide information about how accurately and effectively the program was implemented as it was designed. Another potential benefit of process evaluation is its capacity for documenting operational effectiveness before a program has matured for outcome evaluation.

Process evaluation results can lead to fine-tuning and redesigning a program in the interests of improving a program (Patton, 1990). Bias can result in program implementations from self-selection by targets or from program actions and can compromise not only the success of the intervention but also assessments of outcomes (Rossi & Freeman, 1993). Process evaluation data can provide a means for identifying areas of bias so that such bias may be treated effectively and objectively when conducting an outcomes evaluation. Bias according to Rossi and Freeman refers to the extent that some target population participants are being served, or covered, more densely than others. Three basic sources of information are used in measuring program coverage: program records, surveys of program participants, and community surveys (Rossi & Freeman, 1993). This information provides a means for identifying areas for improving the program as a whole or components of the program where problems with service delivery and bias in participation occur.

Process evaluation is synonymous with program monitoring (Patton, 1990; 2002; Rossi & Freeman, 1993). Four methods should be considered in the design of a process evaluation according to Rossi and Freeman: (1) direct observations, (2) review of service records, (3) interview of service providers, and (4) interview of program participants. Where qualitative information is obtained from staff about features of the program, information can allow for not only identifying staff professional development needs, but also creative approaches to implementing or utilizing program features. Other investigations may also be conducted.

Rossi and Freeman identify three types of investigations that may be conducted using a process evaluation. The first is a description of the program. A useful description of a program estimates a coverage and bias in participation. It may identify the types of services delivered and classify the intensity of services given to participants and the reactions of participants to the services provided. Both quantitative and qualitative data may be used for a description of the program. Descriptions of the program may provide program designers a unique understanding of variables which were either not intended or were not anticipated with respect to implementation of and participation in the program.

The second general type of investigation using process evaluation results is a comparison between sites where a particular program is implemented. This permits an understanding of sources of variability in implementation and outcomes. Sources for obtaining information may include staff, administrators, target populations, and surrounding environment (e.g., competing programs for existing services). Such information may help standardize implementation efforts or offer clues why the program works at some sites and not at others.

The third type of investigation is the level of conformity of the program to its design. This type of investigation is most consistent with measuring the fidelity of a program's implementation and use. Discrepancies may be dealt with by moving the implementation towards the intended program specifications or providing a re-specification of the design itself where adaptations are required for implementation at the local level. Such an investigation provides an opportunity to judge the appropriateness of

outcome evaluations and to determine whether or not a more formative evaluation is required to develop the desired balance between design specification and local implementation requirements.

This type of investigation is consistent with McGraw, et al. (1996) who suggested that when there is a lack of information about what actually occurred during the implementation of an intervention, investigators may not be adequately interpreting study results of an outcome evaluation. Data on the extent or dose of the intervention delivered, the fidelity with which the intervention was carried out, and the presence of competing programs, events, or other confounding influences can be useful for the interpretation of study results (McGraw, Sellers, Stone, Bebachuk, Edmundson, Johnson, Bachman, & Luepker, 1996). According to these authors, and consistent with Rossi and Freeman (1993), process data are useful for describing the program implementation, providing information for quality control and monitoring, and help explain program effects, or outcomes.

Finally, Patton (1990) depicted process evaluations as “elucidating and understanding the internal dynamics of how a program, organization, or relationship operates (p. 95).” He provided several types of questions that are consistent with this approach:

1. What are the things people experience that make the program what it is?
2. What are the strengths and weaknesses of the program?
3. How are clients brought into the program and how do they move through the program once they are participants?

#### 4. What is the nature of the staff-client interactions?

One potential benefit of conducting process evaluations of a program is that it provides outside persons not intimately involved in the day-to-day operations of a program (e.g., funding agencies, public officials, supervising agencies, etc.) to understand how a program operates. Patton uses the terms formative evaluations and process evaluations synonymously. Both are aimed at program improvement and often rely heavily, if not primarily, on process data. In general, process evaluations focus on *how* something happens in a program rather than on the outcomes or results of a program.

#### *Process Evaluations in Education*

Though some process evaluation research has been conducted in education settings or related fields, process evaluation studies conducted on the *Reading First* program could not be found. However, some research and discussion was available which offered insight on the factors which may impact a process evaluation in an educational setting. For example, Fitzgerald and Clark (1976) conducted a process evaluation of inservice training for a school reading program. The overall purpose of the study was to evaluate the effectiveness of the inservice training using teacher self-ratings, director ratings, and monitor ratings as well as any correlations between these ratings. In general, the study found significant discrepancies between the three sources of information. Nonetheless, these authors argued that product and process evaluation are essential to provide information on which to judge the effectiveness of inservice training.

The immediate goal of inservice training is a process change in which the teacher demonstrates both cognitive and behavioral development as a result of inservice training.

Studies investigating process variables have been criticized for not including judgments about changes in teacher performance in the classroom as a result of inservice training. Researchers argue that more information is provided when attitudes about program and self-ratings are used simultaneously with descriptions of organizational process variables (Fitzgerald & Clark, 1976).

Chen and Rossi (1983) contend that teachers in an education organization are notoriously difficult to control and influence because the instructional approaches used by teachers are not directly interdependent on the instructional approaches used by other teachers. It may be common to find variability in the level of fidelity in implementing an educational program; and may also influence teacher variability in the level of participation in a program. It would be important to consider evaluation questions in terms of relative differences between sites; for example, in relation to implementation of program components and participation levels among teachers. Chen and Rossi suggest the implementation of large-scale educational programs relies heavily on the adherence of teacher's behavior to the program's intended design where teachers are depended upon for delivering any component of the program. Because individual schools may be viewed in many respects as independent entities apart from the operations of other schools, and contain personnel who do not have a direct dependence between themselves in the individual school, there is high probability that implementation of large scale education programs will demonstrate modest to low levels of fidelity where specific design features are intended for implementation.

Harachi, Abbott, Catalano, Haggerty, and Fleming (1999) discussed the use of process evaluation studies in the field of education. The 1990s witnessed a greater emphasis on process evaluation of prevention programs in education which attempted to elaborate on mechanisms through which outcomes operate. Examination of program implementation is particularly crucial in the design of efficacy studies of school-based preventative programs according to the authors. Harachi et al, state the extent to which teachers deliver a particular program component as designed is a critical question that needs to be addressed when evaluating outcomes. They argue evaluators must assess program fidelity as well as program outcomes in order to best differentiate the failure or success of implementation as a context for examining the failure or success of a program. Understanding teacher participation levels and perceptions about a program would be an important set of variables when a process evaluation is conducted in education; even if only a partial evaluation of process variables is undertaken.

Considerable variation in implementation of even similar programs is likely to yield differential outcomes. Though process evaluation studies applied to *Reading First* programs is lacking, process evaluation studies have been conducted in other educationally-based areas such as staff training in health education (Helitzer, et al., 2000), substance abuse treatment (Battistich, Schaps, Watson, & Solomon, 1996; Elliott & Mihalic, 2004; Pentz, Trebow, Hansen, MacKinnon, Dwyer, Johnson, Flay, Daniels, & Cormack, 1990), community mental health programs (Durlak & Wells, 1997), drug prevention programs (Goodstadt, 1988; Schaps, Moskowitz, Malvin, & Schaeffer, 1986), and nutrition, (Helitzer, Davis, Gittelsohn, Gohn, Murray, Snyder, & Stecker 1999).

Some researchers have argued that although much information is available for how to conduct a methodologically sound evaluation study, few such studies may be found given the political and financial environments in which a program is often evaluated within (Goodstadt, 1988). Goodstadt, in a review of evaluation research regarding school-based drug education programs in North America, hypothesized variability in evaluation research design and implementation exists due to reservations on the part of administrators and program developers. Reservations may often be due to the commitment of financial resources needed for planning and executing such evaluations. Evaluation research may often cause delays in developing a program while it is being tested in the formative stages of the evaluation. Administrators often assume that implementing evaluation studies requires a specialized evaluator when in fact, according to Goodstadt, specialized evaluators are often only needed for research design and data analysis. Given these observations, considerable variation is to be found among evaluation research studies regarding methodology and implementation of design features.

Durlak and Wells (1997) provided a meta-analytic review of primary prevention mental-health programs. In their review they observed that few studies provided any relevant data regarding how program implementation influenced outcomes. As such, it is apparent that future process evaluation studies will be necessary in order to effectively understand outcome effects, as well as make comparisons between studies.

Schaps, et al. (1986) described a series of seven evaluation studies of school-based drug prevention strategies. The seven strategies consisted of 4 in-service teacher

oriented strategies that focused on classroom management, student attitudes towards school, student self-esteem, and student development of social competencies. Two strategies consisted of academic elective courses for students to enroll which taught skills and provided opportunities for helping peers. None of the six strategies above directly addressed drug use. The final strategy was a drug education course that taught students social skills and provided drug facts and information. The study evaluated the implementation and effects of the strategies listed above. Process data were gathered to monitor implementation. Pre-test and post-test information was used to evaluate outcomes. Process evaluation methods consisted of teacher attendance at each session of training, anonymous teacher ratings of sessions, observed participation in sessions, and documentation of agenda, content, and procedures used in each session. The results of this study found a lack of effect for all strategies, with the exception of partial support for the efficacy of the drug education program. These researchers argued the lack for support for the programs was due to an ineffective theory of drug prevention as opposed to limitations in the design and implementation of their evaluation research.

Battistich et al., (1996) found higher degree of implementation integrity led to greater outcome results in a study examining teacher's implementation of the Child Development Project; a program designed to reduce risk factors and increase protective factors among children. They also found decreases in substance abuse among students in the high and moderate implementation schools while increases were found in comparison populations. Though their findings did not yield a statistically reliable effect given the

decreased sample sizes for the subgroup analyses, their results do suggest the quality of intervention implementation should be considered when evaluating program outcomes.

Pentz et al., (1990) found considerable variability among school-based prevention programs in terms of maintenance of effects, amount of time between provision of the intervention and observation of effects, and the magnitude of the effects found.

Furthermore, such variability was found between studies based on similar methodology and content. Pentz et al., hypothesized variability observed among these prevention evaluation studies may be due to differential implementation of evaluation design and interventions. These researchers offered three approaches for defining quality of implementation: adherence—degree to which the intervention is provided to the experimental group and not the control group; exposure—amount of intervention provided to the target group; and reinvention—degree to which implementation deviates from the designed program standard. Reinvention in particular is of most interest given the unique conditions that often exist at the local level when a program or program component is implemented on a large scale in an educational setting (e.g., state-wide or even county-wide). Furthermore, Pentz found difficulty in measuring the degree to which teachers deviated from implementation design when using self-report surveys. They found 100% of the teachers survey reported they did not deviated substantially from the designed intervention. Those that had deviated to some degree indicated the inclusion of additional materials, discussion, or number of sessions provided.

Given the contextual, political, and cultural constraints of implementing large-scale educational programs, Harachi et al. suggest evaluators look at variables other than

just fidelity. One example is to measure the differences between implementers and non-implementers on variables such as years of teaching experience, degree of self-efficacy, enthusiasm, preparedness, teaching methods compatibility, and principals' encouragement and support. The above variables may also help to identify the conditions under which a program is highly and accurately implemented as designed. These researchers advocate for a complete description and operationalization of the key elements of a program as a first step in designing a process evaluation in education.

Helitzer et al. (1999) utilized process evaluation methods as a formative tool to improve and refine an intervention while being implemented as well as explain and interpret intervention outcomes. These researchers developed a process evaluation plan by outlining a sequence of steps of the intervention that theoretically leads to the desired outcome. In this plan they identified the most salient indicators to measure implementation. The study used the following types of dependent measures: (1) self administered questionnaires that included facilitator characteristics, (2) training evaluation forms completed by facilitators, (3) facilitator check-off lists that documented curriculum-implementation, and (4) observations of group sessions.

### *Qualitative Research Methodology*

#### *Value and Nature of Qualitative Research Approaches*

Any research proposal involving qualitative research methodologies must provide a rationale for engaging in such an enterprise (Marshall & Rossman, 1999). Such a justification involves an understanding of the value and nature of qualitative research.

Patton (1990) reasoned that one of the powerful benefits of qualitative investigations is

that is allows the researcher to understand a particular phenomenon of interest from the points of view of people directly involved in the issue or phenomenon rather than through predetermined views held by the researcher. This rich, detailed characteristic is one of the hallmarks of qualitative designs (Marshall & Rossman, 1999; Maxwell, 2005; Patton, 1990; 2002; Taylor & Bogden, 1998). Qualitative researchers are naturalistic in their approach to inquiry (Patton, 1990; 2002; Taylor & Bogden, 1998). Qualitative studies allow the evaluator to study topics of interest by approaching fieldwork without being limited by predetermined categories of analysis (Patton, 1990). Qualitative research provides a wealth of information about a relatively smaller group of people that although reduces generalizability, increases understanding in greater depth than can be achieved through traditional quantitative approaches.

Qualitative inquiry is not a single thing with a single subject matter (Patton, 1990; 2002). Patton indicated qualitative inquiry builds on several interconnected themes which may be realized in the real world depending on the purpose, situation, questions being asked, and availability of resources. Specifically, he lists 10 such themes: Naturalistic, inductive analysis, holistic perspective, qualitative data, personal contact and insight, dynamic systems, unique case orientation, context sensitivity, empathic neutrality, and design flexibility. Any combination of these can occur and serve as a foundation for conducting a qualitative study. A naturalistic theme focuses on studying the real world situations as they unfold with no attempts to manipulate or otherwise interfere with those situations. An inductive analysis theme refers to an immersion into details and specifics of the data to discover important categories, dimensions, and

interrelations by beginning with exploration first rather than testing theoretically derived hypotheses. A holistic perspective involves seeing the phenomenon as a whole whereby it is greater than the sum of its parts given its complexity as a system. Personal contact and insight refers to direct contact with people, situations, and the phenomenon of interest whereby the researcher brings their personal experiences and insights as an important part of the inquiry and critical to understanding the phenomenon. A dynamic system theme involves paying attention to processes and assuming change is constant and ongoing whether the focus is on an individual or an entire culture. Unique case orientation themes assume each case is special and the focus is on capturing details in individual cases and then engaging in cross-case analyses based on individual case studies. Context sensitivity refers to placing findings within a social, historical, and/or temporal context. The theme of empathic neutrality assumes complete objectivity is impossible and pure subjectivity undermines credibility. Thus, the researcher's role is not in proving anything, advocating any particulars, or otherwise advancing any personal agenda. Finally, Design flexibility refers to the openness for adapting the inquiry as understanding deepens and/or situations change by avoiding rigid designs that eliminate responsiveness, but rather pursues new paths of discovery as they emerge.

Qualitative methods are particularly suited for studies focusing on process variables involved in a program or organization (Marshall & Rossman, 1999; Maxwell, 2005; Patton, 1990; 2002). Such investigations allow for an understanding of (a) implementation fidelity, (b) program operations and innovations where specific guidelines or rules are not provided in the implementation plan, (c) attitudes, beliefs, and

perceptions held by those serviced by a program or who participate in the operation of a program—this is particularly of interest where participants in the program’s operation are depended upon for delivering the appropriate levels of the intervention/service to clients, and (d) program components or issues that require modification for overall program improvement. As Patton suggests, a common activity for all can result in drastically different outcomes depending on how the activity is experienced, what the unique needs are of each individual, and which parts of the activity were found to be most salient for individuals. Qualitative studies are further well suited for understanding individual client outcomes (Marshall & Rossman, 1999; Patton, 1990; 2002; Taylor & Bogden, 1998), programs comparisons (Patton, 1990; 2002), evaluability assessment – the degree to which a particular program is ready for systematic quantitative evaluation (Patton, 1990; 2002), and quality assurance (Marshall & Rossman, 1999; Patton, 1990; 2002; Taylor & Bogden, 1998).

In their book, *The Discovery of Grounded Theory*, Glaser and Strauss (1967) argue qualitative research should focus its efforts to developing social theory and concepts. Their Grounded Theory approach was designed to enable researchers to find plausible support for theories through discovery from data rather than from a prior assumptions, other research, or existing theoretical frameworks (Taylor & Bogden, 1998). The Grounded Theory approach begins with collection of data about a particular phenomenon from which themes, concepts and ideas are based upon. Reviews of the themes, concepts and ideas are conducted in addition to collection of other data for comparisons and refinement of ideas. Convergence of information builds towards the

establishment of a theory which “fits” the data. According to Glasser and Strauss (1967), the main criteria in evaluating theories are if they “fit” and “work.”

Other researchers have called into question this emphasis on focus for pursuing qualitative investigations (Patton, 1990; Seale, 1999; Taylor & Bogden, 1998). Seale (1999) for example differentiated between social or cultural commentary against social research. Patton (1990) allows for a more pragmatic approach whereby qualitative studies may be conducted to describe human action, beliefs, attitudes, and understandings regarding a program. More specifically, Patton wrote, “...there is a practical side to qualitative methods that simply involves asking open-ended questions of people and observing matters of interest in real-world settings in order to solve problems, improve programs, or develop policies (p. 89).” His view on this issue, as he further argues, is like saying one may utilize statistical methods in a relatively straightforward way and not necessarily include an exhaustive review of logical-positivism. “While these intellectual, philosophical, and theoretical traditions have greatly influenced the debate about the value and legitimacy of qualitative inquiry, it is not necessary, in my opinion, to swear vows of allegiance to any single epistemological perspective to use qualitative methods (Patton, 1990; p. 89).”

#### *Purpose and Research Questions*

A qualitative inquiry must provide a detailed purpose and justification for using a qualitative approach to investigating a particular phenomenon (Marshall & Rossman, 1999; Patton, 1990). Marshall and Rossman provide several general justifications when conducting such a study. One in particular bears specific mention as it relates to the topic

chosen in the present study: Research on informal and unstructured linkages and processes in an organization. This justification is consistent with Patton's (1990) descriptions of formative evaluations, or process evaluations whereby understanding the specific actions as they occur in the natural setting can assist in identifying potential areas for improvement. One of the potential purposes of engaging in a qualitative approach is that this approach allows for a collection of rich and detailed information about the beliefs, attitudes, or general understandings from the perspective of the individuals participating. As such, qualitative studies tend to focus on three types of questions according to Maxwell (2005): (a) understanding of the meaning of events and activities to the people involved; (b) understanding the influence of the physical and social context on these events, and (c) process questions by which these events and activities occur.

Similarly, Marshall and Rossman (1999) describe three purposes of qualitative research: (a) exploration of phenomenon not well understood in order to identify important categories of meaning to generate hypotheses for future research, (b) explanation of patterns related to a phenomenon in order to identify plausible relationships shaping the phenomenon, and (c) description and documentation of the phenomenon of interest in order to identify the salient actions, events, beliefs, attitudes, or social structures and process occurring in a given phenomenon or topic of interest. Questions which focus on differences of variance are considered appropriate for quantitative studies (Maxwell, 2005; Patton, 1990). However, questions which have as their goal an understanding of what is actually taking place in the field is best supported by a qualitative approach (Maxwell, 2005; Patton, 1990).

## *Qualitative Research Methods*

*Researcher's Biography and Role.* Researchers recommend a qualitative study must be specific and upfront of the researcher's role in the study. According to Marshall and Rossman (1999), a qualitative study must provide in its proposal a clear description of the researcher's role in the study, level of reciprocity provided, plans for entering the organization being studied, a description of the researcher's assumptions and potential biases, and consideration of any ethical concerns. Patton (1990) provided five dimensions for consideration when designing a qualitative study. These dimensions each lie on a continuum and are defined as follows:

1. Role of the Evaluator-Observer – the extent to which the researcher participates in the daily activities of the participants in their natural setting (full – partial – onlooker).
2. Portrayal of the Evaluator Role to Others – the extent to which participants know that observations are being made and who the observer is (overt – partial knowledge – covert).
3. Portrayal of the Purpose of the Evaluation to Others – the extent to which participants are informed of the purpose of the research study (full – partial – covert – false explanation).
4. Duration of the Evaluation Observations (single observation with limited time – long-term multiple observations).
5. Focus of the Observations (narrow/single element or component in a program – broad focus/holistic view of the entire program and all its elements).

These five dimensions are used by Patton to describe the variation in approaches to observations which can occur in qualitative studies. They provide a framework for the researcher to define the parameters of a study and review how the evaluation is proceeding. As Patton states, “It is not possible to observe everything (p. 216).” By focusing the study along very specific dimensions the researcher can find organization and logic while attempting to observe the complex reality of a given situation or program.

Marshall and Rossman (1999) also suggest any qualitative study must contain a researcher biography which specifically highlights any *a priori* expectations, beliefs, and/or biases the researcher may have about conducting the study. The purpose of providing such a biography is to provide greater credibility to the reader in providing up front the experiences, biases, expectations, and/or beliefs about the topic being investigated. Consistent with Patton’s concept theme of empathic neutrality, one cannot present themselves as a clean slate. Qualitative research does not require the researcher to remove such subjectivity. Rather, the goal is to acknowledge subjectivity and understand how it influences the study. This is discussed more in the following pages concerning validity and reliability constructs in qualitative studies.

*Selection Methods.* Researchers have indicated there are no formal rules for sample size in qualitative studies (Marshall & Rossman, 1999; Patton, 1990; Taylor & Bogdan, 1998). The size of the sample depends on what the researcher wants to know, the purpose of the study, what information is intended to be useful to stakeholders, what will have credibility, and what can be done with available time and resources. The size of the sample chosen for a qualitative study must be large enough to capture most or all

of the perceptions that might be important for inclusion in the study. A sample size that is too narrow runs the risk of narrowing the scope or range of information available within a population. A sample size that is very large has less risk regarding threats to validity; however, resources may be limited for designing studies with very large samples. Marshall and Rossman (1999) argue that the number chosen for a sample size must be thoroughly justified in any research proposal involving qualitative research. Finding a balance between breadth and depth of a study is essential when conducting a qualitative study. A method is needed for determining a reasonable starting point that finds balance between available resources and assuring a majority of potential perceptions are captured within the study. Patton recommends researchers provide a minimum sample level for selection. The researcher may add to this number as the study unfolds.

Researchers have suggested recording consecutive individual interviews until a saturation point is reached – that is, a point in which the information obtained is no longer new (Fossey, Harvey, McDermott, & Davidson, 2002; Patton, 1990; 2002). However, according to DePaulo (2000), such an approach is not solidly grounded and may not tell the researcher in advance the optimal qualitative sample size needed which maximizes capture of new and relevant information while also minimizing unnecessary resource expenditures. Starting with a sample size of 30 may be considered a reasonable starting point when random sampling procedures are employed (DePaulo, 2000; Griffin & Hauser, 1993).

Specifically, DePaulo reasoned that if the incidence rate of an assumed sub-population (e.g., consumers who are dissatisfied with a product) is 1 in 10, then the chance of randomly selecting a person with a *satisfactory* perception is 0.9. To determine the chance of finding the same type of participant with a satisfactory perception a second time in a row, one would multiply 0.9 to the second power. Therefore, the probability of selecting 10 satisfied participants in a row would be 0.9 to the tenth power, or 0.35. DePaulo reasoned there would be a 35 percent probability that a sample of 10 would have missed participants who were dissatisfied assuming an incidence rate of 1 in 10. DePaulo developed a table for researchers to use as an initial guide to minimize the risk for sampling error by repeating the power calculations for various incidences and sample sizes. The table is then useful for selecting a sample size that is within the resource limits while also reducing the risk of sampling error. His calculations suggested 30 participants with an assumed incidence of 10% would have a 95% chance of capturing all potential perceptions in the population sampled. This starting point is consistent with Griffin and Hauser (1993) who found fewer new perceptions with each additional interview conducted beyond 30 individual interviews in a study involving consumer use of products. Again, this assumes a population incidence rate of 10%. Conversely if one assumes an incidence rate of 33% then according to the table, there would be about a 98% chance of capturing all relevant perspectives.

Following selection of the number of participants for inclusion in a qualitative study, the researcher must consider the appropriate sampling method for selecting participants. Some researchers have provided comprehensive lists which describe

various types of methods for sampling (Miles & Huberman, 1994; Patton, 1990). Patton in particular provided a list of 16 methods. The reader is referred to Patton for a complete description of the strengths and weaknesses of each method which is beyond the scope and purpose of this review. A few methods do deserve some attention: Criterion sampling, random purposeful sampling, and convenience sampling. Criterion sampling procedures involve picking all cases that meet some criterion. According to Patton, this approach is common in quality assurance studies. This sampling method can be applied to identify cases from questionnaires or tests for in-depth interviews. Random purposeful sampling adds credibility to a study when the potential pool of participants is too large for one to handle. It also reduces judgment within a purposeful category. The weakness of this approach is that it does not allow for generalization or representativeness. This may not be an issue for studies where the focus involves a descriptive purpose for engaging in the qualitative study, or generating potential hypotheses for future research. Convenience sampling methods serve as the poorest rationale for sampling participants (Maxwell, 2005; Patton, 1990; 2002). However, this approach is focused on minimizing costs and reducing time.

*Data Collection and Analysis.* Great care and consideration must be involved when designing data analysis procedures in a qualitative study. Researchers have noted that the data analysis part of any qualitative study is often the weakest part of the study (Maxwell, 2005). Taylor and Bogden (1998) described qualitative data analysis as, “the most difficult aspect of qualitative research to teach or communicate to others (p. 140).” According to Miles and Huberman (1994), there are few agreed upon rules for drawing

conclusions and verifying the sturdiness of the conclusions. What is agreed upon in the research literature is the need to conduct analyses of the data concurrently with data collection, and provide rich detailed information concerning the physical and social contextual variables; in addition to the researcher's thought processes throughout the analysis of results. Patton (1990) argued the researcher has an obligation to, "monitor and report their own analytical procedures and processes as fully and truthfully as possible," regardless of the analysis procedures employed (p. 372)." That is, the researcher will not only report on the findings of the study, but also on the analytical process.

Marshall and Rossman (1999) described typical qualitative analysis procedures as having six phases: (a) organizing the data; (b) generating categories, themes, and patterns; (c) coding the data; (d) testing the emergent understandings; (e) searching for alternative explanations; and (f) writing the report. Maxwell (2005) conceptualized analytic options into three broad categories: (a) memos; (b) categorization strategies; and (c) connecting strategies. Categorization strategies include the popular procedure of coding the data. Categorization strategies may be classified into three types: organizational, substantive, and theoretical. Organizational categorization involves predetermined identification of broad topics or categories which may be later used as "bins" for presenting the results of the study. Substantive categorization primarily involves a description of the participants' concepts or beliefs. This process involves inductive analysis, and the outcomes of this process may be used to facilitate a more general theory of what is going on in the study, but does not necessarily depend on the

theory being explored. Theoretical categorization approaches typically represents the researcher's concepts rather than denoting the participants' own concepts and the topics developed may be either derived from existing theories or from an inductively developed theory.

Taylor and Bogden (1998) suggested "discovery" is the first step in the analysis of qualitative data. This ongoing process of discovery involves identifying themes and developing concepts and propositions. Given the sheer volume of information (measured in pages of text) it is essential that data analysis procedures be ongoing throughout the data collection process. The cycle of collection and analysis allows the researcher to track early themes and concepts, either intended or unintended, which offer the researcher a means of refining, highlighting, or otherwise focusing on such themes and concepts in subsequent data collection sessions. Coding the data is the second step for Taylor and Bogden. This process usually occurs after the data have all been collected. These researchers caution the reader about any delay in coding the data after it has been collected since the greater the delay the greater the difficulty may occur with going back to informants to clarify any points or tie up loose ends. Researchers according to Taylor and Bogden may even maintain casual contact with informants during the analysis phases and/or have informants read draft reports as a check on interpretations. Finally, the third step involves attempting to discount findings, or understand the data in the context in which they were collected. Field notes are essential for undertaking this step (Patton, 1990). This final step is consistent with Patton's recommendation to researchers that the analytical process be communicated clearly and in great detail when writing the results of

the study. Thus, making sense of qualitative data must include the researcher's role, the context in which the data was obtained, informal observations made before, during, and following interviews or focus groups, and the researcher's thinking process for making specific interpretations and the evidence used to make those interpretations.

More recently computerized software programs are being utilized for the analysis of qualitative data. According to some researchers there is a seductive aspect about this approach to analysis (Fielding, 1993; Taylor & Bogden, 1998). According to Fielding (1993), there is a "perceived danger of superficial analysis produced by slavishly following a mechanical set of procedures (p. 3)." Fielding argues that most software may have an implicit conventional grounded theory approach; however such programs may not be valuable for researchers interested in hermeneutic approaches, ethnomethodology, conversation analysis, or holistic analysis. Nonetheless, computerized analysis can provide a measure of reliability in the coding of data for identifying themes or concepts towards interpretation (Seale, 1999; Taylor & Bogden, 1998)

Patton (1990; 2002) referred to two analytical approaches as "case analysis" and "cross-case analysis". Patton suggested case studies are appropriate when the researcher is interested in variation in individuals as the primary focus. Cross-case analyses involve "grouping together answers from different people to common questions or analyzing different perspectives on central issues (p. 376)." When an interview guide is used as the tool for collection of interview data, people can be grouped by topics from the guide. It is important that the researcher understands upfront that the relevant data won't

necessarily be found in the same place for each interview given the natural conversational flow required of an interview process.

*Reliability and Validity.* Any discussion of data analysis procedures must include a discussion of validity and reliability (Maxwell, 2005; Patton, 1990; 2002; Seale, 1999). The validity and reliability of qualitative data depend to a great extent on the methodological skill, sensitivity, and integrity of the researcher (Patton, 1990; 2002). Unlike traditional quantitative methodologies which rely upon well designed instruments and statistical procedures for ensuring the reliability and validity of the study, qualitative studies must make explicit the use of any procedures for minimizing threats to validity and ensuring internal reliability since the researcher is the research instrument (Marshall & Rossman, 1999; Maxwell, 2005; Patton, 1990; Seale, 1999).

According to Seale (1999) attempts at replicability of qualitative studies is rarely found and where found are exercises to promote artificial consensus. In other words, according to Seale, different people are likely to have different accounts of the world. This view seems consistent among other researchers of qualitative studies. Taken as a whole, the research literature suggests the concepts of reliability and validity are not readily applicable to a purely qualitative study. Nonetheless these concepts are consistent within a qualitative approach when conceptualized as determining the “trustworthiness” of the researcher’s interpretations of the information. Therefore, one obligation the researcher has in order to demonstrate a trustworthiness is to “facilitate the expression of these accounts” (Seale, p. 42) so that others may make their own analyses to compare with the researcher’s.

However, this may not necessarily provide trustworthiness if such facilitation is viewed as selections of possible versions. To make sense of this, Seale recommends researchers conceptualize the reliability aspects of a qualitative study as having either external or internal reliability. External reliability is a demanding process referring to the replication of findings in conducting the same study again. In qualitative research, the sheer variety of different points of view can often lead to different outcomes in a replicated study. Thus, “the expectation of complete replication is a somewhat unrealistic demand (p. 42).” Internal reliability is much more obtainable in qualitative studies and there are a variety of strategies that may be used to demonstrate a qualitative studies’ internal reliability. Internal reliability can refer to the extent to which different researchers identify similar constructs (Seale, 1999). Computers may also be used to facilitate the finding of coding errors or discrepancies according to Seale. This basic procedure typically involves a use of inter-rater checks on the coding process.

Lincoln and Guba (1985) reasoned that the conventional terms of validity and reliability may be inappropriate for qualitative work if a “naturalistic” approach is used which rejects a cause and effect view of the world; typically characteristic in a positivist view of the world. They conceptualized reliability and validity within a qualitative study as one of “trustworthiness” being established. They identified four additional ways a researcher can establish trust in their research project: (a) internal validity or “credibility”; (b) external validity or “transferability”; (c) reliability or “dependability”; and (d) objectivity or “confirmability”. Credibility can be achieved by engaging in long durations in the field with participants, engaging in multiple and persistent observations,

engaging in triangulation exercises, and exposing the research report to others and/or searching for alternative interpretations. Transferability is not intended to be achieved by use of random sampling or probabilistic reasoning. Rather the researcher provides a detailed, rich description of the setting studied so that readers are given sufficient information to be able to judge the applicability of findings to other settings which they know. Dependability can be achieved by engaging in “auditing” which consists of the researchers’ documentation of data, methods, and decisions made during a project, as well as its end product. Auditing is also useful in establishing confirmability. Auditing, according to Lincoln and Guba is also conceptualized as an exercise in “reflexivity” which involves the provision of a methodologically self-critical account of how the research was done, and can also involve triangulation exercises.

#### *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*

Early literacy skills have been a focused area of research in the last couple of decades. More importantly, advances have been made with regards to intervention and assessment of early literacy skills in the last decade. Adams’ (1990) book entitled *Beginning to read: Thinking and learning about print* was written based on a comprehensive review of the research literature on reading. Adams concluded a student’s trajectory of reading development was increasingly highly resistant to interventions if reading difficulties persisted beyond the first grade. However, her research suggested that a student’s trajectory was sensitive to change during the pre-kindergarten through first grade years. Johnston and Allington’s (1991) review of remedial reading intervention studies was consistent with Adams’ findings in which they

too concluded remedial reading instruction has not been very effective, in general, at making children more literate. Prevention has therefore gained more attention as the optimal course of action in developing students' reading literacy.

In 1996, Good & Kaminski published a manuscript entitled *Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills*. This article provided a case study to illustrate the use of DIBELS within a problem-solving model of educational decision-making. The purpose of the article was to demonstrate how to use DIBELS to develop local norms, monitor progress of student performance, and evaluate effects of interventions on an individual basis. It emphasized a link between teaching and assessment in which assessment is most useful when it is directly linked to the curriculum. It also emphasized a need for early intervention for at-risk learners in reading before the end of first grade. The goals of the decision-making model using DIBELS are to (1) prevent reading problems by ensuring that students have early literacy skills, and (2) proactively engage in instructional modifications to minimize the magnitude of reading problems for students who are having trouble developing early literacy skills. In short, DIBELS were developed as a prevention approach to reading improvement.

DIBELS are not intended to be the sole form of assessment used with young children (Kaminski et al., 2008). Convergent information from multiple sources and multiple assessment procedures is desirable especially in high stakes educational decisions. The methods used in the DIBELS measures do not suggest the methods of instruction. DIBELS were developed to serve as indicators of critical early literacy skills.

The measures link most directly to three of the five “big ideas” or essential skill areas identified in the reading research literature (National Reading Panel, 2000). Those three skill areas are (1) phonological awareness, (2) alphabetic principle, and (3) fluency with connected text. Phonological awareness refers to the explicit awareness of the *sound* structure of language, including the ability to orally manipulate sound units smaller than words. This skill area includes rhyming words, blending phonemes, segmentation of phonemes and syllables, and deletion of phonemes and syllables. Phonemic awareness is not the same thing as *phonics*, which is the pattern of letter-sound correspondences in *written* language. The alphabetic principle refers to a child’s awareness of letter-sound correspondences. Fluency refers to a child’s oral reading rate at a particular instructional level within a standardized time frame.

As part of the early development of DIBELS, Kaminski & Good (1996) examined the reliability, validity, and sensitivity of experimental measures to assess areas of early literacy. The measures were phoneme segmentation fluency, letter naming fluency, and picture naming fluency. The measures were designed for repeated use to identify children with difficulty acquiring basic early literacy skills. The measures were tested on 37 kindergarten students who could not read and 41 first grade students who could read. All 78 students were drawn from one elementary school in the Pacific Northwest. Both the kindergarten and first grade groups were divided randomly into two groups, a monitored and non-monitored group. Students in the monitored groups were administered the measures two times a week for a period of 9 weeks while students in the

non-monitored groups were tested with the measures only at the beginning and at the end of the 9 week period. The study was conducted in three phases.

In phase one all students were tested on the McCarthy Scales of Children's Abilities and the three DIBELS measures. Their teachers completed the Rhode Island Pupil Identification Scale and a teacher rating scale in order to further evaluate students' classroom behaviors. Kindergarten teachers also administered the Metropolitan Readiness Test, which assesses reading and math readiness skills. The first grade teachers administered the Stanford Diagnostic Reading test and CBM reading as a measure of reading readiness. Phase two consisted of the 9-week progress monitoring of students assigned to the monitoring groups. Phase three involved all measures except the McCarthy. Results were summarized in three ways: point, level, and slope of scores.

Reliability estimates on the three DIBELS measures for the kindergarten students ranged from .77-.93 (point), .97 to .99 (level), and .62-.88 (slope). The reliability estimates for the first graders were .60-.83 (point), .83-.95 (level), and .00-.36 (slope). The authors concluded moderate to high reliability for kindergarten students on the use of the DIBELS measures. The first grade reliability coefficients were low to moderate. Criterion validity was examined by correlating the DIBELS estimates (point, level, and slope) with the criterion tests used in the first phase of the study. For the kindergarten students, significant positive correlations were found for all point and level estimates with all criterion measures (range = .43 to .90,  $p < .01$ ). All but one of the twelve correlations for slope estimates for kindergarten students were not significant. Furthermore, the correlations for slope on the letter naming DIBELS measure were

negative with the first one (correlated with the McCarthy) being  $-.59$  ( $p < .05$ ). Only 9 of the 45 correlations for the first grade students on point, level, and slope estimates were significant positive correlations with two of them significant at the  $.05$  level. (range =  $.42 - .74$ ,  $p < .01$ ;  $.32 - .47$ ,  $p < .05$ ). Analysis of sensitivity yielded acceptable evidence that children's performances on the measures may change substantially from the end of the kindergarten to the end of the first grade years.

The phoneme segmentation measures proved to be a more reliable, valid, and sensitive measure than the letter naming and picture naming measures. Furthermore, slope estimates for the kindergarten and first grade groups were very low and suggested little change during the 9 week monitoring period. The authors concluded that the failure of the picture and letter naming fluency tasks on slope estimates may indicate that these two measures are not very sensitive to performance change over a 9 week period. They further hypothesized that such little evidence of change may be a result of little instruction provided for those skills in the classroom during the 9-week monitoring period. Authors correctly concluded that the sample size of the study was limited and cautioned the reader about generalizing their findings while calling the results of the study preliminary.

According to Kaminski and Good (1998) reading is a cultural imperative in today's information-based society. The authors argue that DIBELS provide a significant value because it provides a cost-effective, time-efficient approach to (1) identifying students who require early literacy skills interventions beyond the general education curriculum, (2) providing formative evaluation regarding intervention effectiveness for

individual students, and (3) determining when interventions have successfully reduced the risk of reading failure by remediation of early literacy skills deficits. The key to the success of DIBELS is its emphasis on formative assessment. By providing repeated measurements of a student's progress over time, teachers are best equipped to make critical instructional decisions that ensure successful progress towards goals. Kaminski and Good also discussed the differences between Curriculum-based Measurement (CBM) Reading and DIBELS. Though the authors list several differences between DIBELS and CBM-Reading which are not directly pertinent to this study, two aspects deserve attention.

Kaminski and Good indicated variability in performance is a natural characteristic of all people. The authors argue CBM, in general, displays a moderate level of variability compared to DIBELS, which may be characterized as having high variability. This variability may be associated with the nature of the skills being assessed (e.g., emerging early literacy skills as opposed to oral reading fluency). As a result, more data points may be required when using DIBELS in order to obtain the same level of confidence given the sensitivity and age of the population being served. Also, CBM has a long duration with reliable use over several years while DIBELS have a short duration with many measures having floor effects and reaching ceiling effects for different time periods with different measures. The authors characterize DIBELS as a necessary but insufficient set of skills towards life skills. Thus it is best to consider DIBELS as an assessment tool for students acquiring early literacy skills only.

Good, Simmons, & Smith (1998) provided a rationale for early and intensive literacy intervention by reviewing three converging areas of research in early literacy and reading acquisition. The authors also provided mechanisms to enhance early literacy development through the strategic and timely linkage of assessment and intervention. The three areas of converging research are essential skills to teach for successful literacy development, research-based instructional designs for teaching essential literacy skills, and utilizing assessment to make curricular and instructional changes in the classroom. The authors highlight a consistent finding that children who experience early reading difficulties are likely to continue to experience later difficulties. However, much of these conclusions come from correlational studies that do not communicate sufficiently the scope of the issue and problem.

CBM-Reading may be used in order to directly monitor a child's reading trajectory. However, such assessment tools may not allow for early detection of difficulties in acquiring early literacy skills which provide the foundation for later reading fluency and comprehension. In this context, Good et al, describe the "Mathew Effect" (Stanovich, 1986) in which initial skills lead to faster rates of acquisition of subsequent skills for those students with high skills and slower acquisition for students with lower initial skills. The authors argue the differences in developmental reading trajectories may be due to a series of reading experiences and activities that began with difficulty in early literacy skills, combined with minimal exposure to print, and further build to lowered motivation and desire to read. They highlight research that shows good readers may have as much as twice the exposure to the amount of vocabulary experienced by poor readers

prior to entering into the kindergarten year. The authors argue low initial skills and low slope in trajectories of reading development lead to a near impossibility for “catching up” to their same-age peers beyond the first grade.

Good, et al. (1998) point to a converging body of research that highlights a number of different areas, or big ideas for early literacy programs. The first of these is phonological awareness, or an individual’s sensitivity and awareness of the sound structure of one’s language. Phonological awareness includes skills such as segmenting, blending, and naming letters and sounds. The second area is an understanding of the alphabetic principle, which refers to an individual’s mapping of print to speech and establishing an association between a letter’s sound and its written form. Phonological recoding is the third area that describes an individual’s ability to recognize the relationship between words, syllables, phonemes, and letter sounds in order to pronounce an unknown written word or to spell a spoken word. The development of this skill is progressive. The easiest form is when a child is provided with a spoken word and asked to match it with three distinctly different written words. Later, children use letter-sound correspondences and their positions in sequences to spell and read words. With redundancy and practice, children learn to recognize words more efficiently. The fourth area concerns accuracy and fluency with connected text. Poor word recognition skills inhibit an individual’s ability to comprehend text. By making these four big ideas explicit components of an early literacy program, early reading acquisition skills may be enhanced for students and provide a solid foundation for continued reading development (e.g., reading comprehension).

How to teach the above skills is also of critical importance. In demonstrating the link between what to teach and how to teach it, Good et al, drew upon the work developed by Smith, Simmons, and Kame'enui (1998) who synthesized the results of 25 intervention studies conducted from 1985 to 1996. Good et al., summarized the five features for teaching phonological awareness. The first is to provide instruction at the phoneme level. The second is to scaffold tasks and examples according to a range of linguistic complexity. Next, explicitly model phonological awareness skills prior to student practice and provide students with generous opportunities to produce isolated sounds orally during practice. Fourth, provide systematic and strategic instruction for identifying sounds in words, blending and segmenting, and culminate with integration of phonological awareness and letter-sound correspondence instruction. Finally, use concrete materials to represent sounds. Good et al, describe the use of DIBELS as a means of evaluating the acquisition of early reading skills and using DIBELS data to make instructional decisions. The authors address the limitations of traditional reading assessments in terms of poor utility towards intervention design and poor utility for formative assessment or progress monitoring.

#### *Reliability and Validity of DIBELS*

Some research and discussion exists in the literature which demonstrates DIBELS utility for early detection of student difficulties in acquiring early literacy skills (Barger, 2003; Buck & Torgesen, 2003; Elliott, et al., 2001; Good, et al., 2001; Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002; Kaminski et al., 2008; Shaw & Shaw, 2002). Elliott, et al. (2001) evaluated four of the DIBELS measures in identifying kindergarten

students who are at-risk for reading failure. These measures were letter naming fluency, sound naming fluency, initial phoneme ability, and phoneme segmentation. The authors point out that the original version of DIBELS was a set of 10 brief measures conceptualized as downward extensions of CBM reading probes. The 10 original measures were: Story Retell, Picture Description, Picture Naming Fluency, Letter Naming Fluency, Sound Naming Fluency, Rhyming Fluency, Blending Fluency, Onset Recognition, and Phoneme Segmentation Fluency. Elliott et al. evaluated two of the three measures evaluated by Kaminski and Good (1996); Letter Naming Fluency and Phoneme Segmentation Ability. The authors also evaluated the reliability and validity of using the Sound Naming Fluency and Initial Sound Fluency measures. The purpose of the study was also to extend the results of Kaminski and Good by including a larger sample size of participants.

The study involved 75 kindergarten students from four classrooms in three elementary schools in a Midwestern city. Although the sample was small, the group was representative of local and national populations in gender percentages, percentage of minorities as a whole, and percentage of population on free or reduced lunch status in school. Five assessments were provided to students: 4 DIBELS measures, the Broad Reading and Skills Clusters on the *Woodcock-Johnson Psychoeducational Achievement Battery—Revised (WJ-R)*, the *Test of Phonological Awareness—Kindergarten form (TOPA)*, and the *Kaufman Brief Intelligence Test (K-BIT)*. School staff earlier in the school year completed the *Developing Skills Checklist (DSC)*. The DIBELS measures were the predictor measures. The WJ-R, TOPA, DSC, and the teacher-rating

questionnaire were used as criterion measures. The K-BIT scores were used to control for differences in ability in the regression analyses.

Inter-rater reliability estimates and coefficients of stability and equivalence for three of the DIBELS measures ranged from .80 - .90 (Letter Naming Fluency, Sound Naming Fluency, and Phoneme Segmentation Ability). The authors concluded that the magnitude of coefficients of stability/equivalence indicated most of the variance in students' scores may be attributable to true differences in abilities measured than to errors in measurement. Concurrent validity of the DIBELS measures were computed by measuring the correlation between level estimates on the DIBELS measures (average scores over repeated administrations) and each of the criterion measures used. The strongest correlations were found between the DIBELS Level estimates and the Skills cluster of the WJ-R and Developing Skills Checklist.

The authors reported 35-40% of the variance in scores on the two achievement measures was explained by correlations between the DIBELS measures and the criterion achievement measures. Correlations between the DIBELS total and the WJ-Achievement Battery were .62 (Broad Reading cluster), .81 (Skills Cluster), and .67 (Letter-Word Identify). Correlations between the DIBELS and the Teacher Rating Questionnaire-prereading, *Developing Skills Checklist*, and *Test of Phonological Awareness (TOPA)* were .67, .74, and .69, respectively. The authors concluded that the DIBELS measures explained at least 16% of the variance and, in most cases, between 30 and 40% of the variance in students' scores when the achievement measures were used as criteria.

Hierarchical regression analyses were computed to further understand the relationship between the DIBELS measures and the achievement measures used in this study. Controlling for ability level and phonemic awareness ability, the authors concluded the DIBELS measures explained a significant proportion of the remaining variance in students' scores on the three achievement measures. The authors concluded the DIBELS measures explained a larger portion of the variance in achievement scores than did either the Kaufman Brief Intelligence Test (K-BIT) or the Test of Phonological Awareness (TOPA). Finally, the authors concluded the Letter Naming Fluency measure of DIBELS was the single best predictor of performance of kindergarten achievement scores on the Broad Reading and Skills clusters of the Woodcock-Johnson Psychoeducational Battery-Revised (WJ-R) and of teachers' ratings of student reading status.

Hintze, Ryan, and Stoner (2003) also analyzed the technical aspects of DIBELS by examining the concurrent validity and diagnostic accuracy against the *Comprehensive Test of Phonological Processing* (CTOPP). The CTOPP is comprised of 7 subtests (Ellison, Rapid Color Naming, Blending Words, Sound Matching, Rapid Object Naming, Memory for Digits, and Non-Word Repetition). The "Ellison" subtest requires a student to repeat a verbally presented stimulus word while omitting a sound. Both rapid subtests are timed measures. Standard scores for each subtest and three composite scores are obtained on the CTOPP. The three composite scores are Phonological Awareness Composite, Phonological Memory Composite, and the Rapid Naming Composite. Subtest Coefficient Alphas on the CTOPP ranged from .74-.93. Internal consistency for

the three composite scores ranged from .84-.97 on the 5-yr old measure, and a range of .81-.96 on the 6-yr old measure. Subtest test-retest reliability ranged from .74-.97. Test-retest scores for the three coefficients ranged from .70-.92. Predictive validity scores with the Woodcock Reading Master Test-Revised ranged from .42-.71.

The study involved 86 kindergarten students in the winter of their kindergarten year. Correlation patterns between the two measures and decision accuracy studies were conducted to determine the appropriate cut-off scores for DIBELS. Students were given both the CTOPP and the DIBELS measures during the same 20-minute session with the sequence of presentation counterbalanced and a short break offered between measures.

Results found DIBELS strongly correlated with most subtest and composite scores of the CTOPP. Results indicate the Initial Sound Fluency and Phoneme Segmentation Fluency measures of DIBELS correlated the strongest with the four CTOPP subtests which measure phonological awareness and memory and less strongly with CTOPP subtests which required rapid naming and memory for digits. Results also indicated the Letter Naming Fluency measure of DIBELS correlated strongly with CTOPP subtests and composite scores that measure phonological awareness and memory as well as rapid naming abilities. Authors conclude these data provide support for LNF as an early indicator of phonological development.

Results of the diagnostic accuracy analysis involved examination of cutoff scores on the basis of sensitivity (accuracy in identification of students at-risk between DIBELS and CTOPP), specificity (accuracy in identifying students who do not present a problem as identified by CTOPP scores), and predictive power (number of false positives and

false negatives in comparison to CTOPP results). Earlier analyses used the cutoff scores for the Initial Sound Fluency and Phoneme Segmentation Fluency measures as suggested by the DIBELS developers. Results indicated the sole use of ISF and the PSF scores on the DIBELS are likely to over-identify students as having a weakness in phonological awareness skills in comparison to CTOPP composite scores. The authors conducted follow up analyses by conducting a series of Receiver Operating Characteristic (ROC) curves to determine appropriate cutoff scores for ISF, LNF, and PSF measures that balance both sensitivity and specificity. Cutoff scores of 15 for ISF, and 25 for LNF were found to provide an appropriate balance between sensitivity and specificity against the CTOPP Phonemic Awareness Composite (PAC) score. The PSF score failed to predict Phonemic Awareness Composite scores over a range of cutoff scores while moderate to high levels of sensitivity were observed when using cutoff scores ranging between 20 and 34. Thus, where the PSF accurately predicts students who exhibit phonological awareness problems on the CTOPP, it does so at the expense of a moderate number of false-positives.

In 2002, Good, et al. published a technical report providing practitioners with updated cutoff scores for determining student need for intensive, strategic or benchmark instruction. The authors indicate that the adjustments were made in order to balance sensitivity and specificity. The reader is referred to their paper for specific criteria found for each measure and for each grade level. In general, however, it is important to point out that the DIBELS is not a nationally norm-referenced measure. It was never designed to be a nationally norm-referenced measure. The flexibility it offers to local education

agencies is that each agency can develop their own grade level norms in order to best meet the educational needs of its students. As such, states which have implemented DIBELS (e.g., Florida) may provide a different set of norms than those offered by Good et al. (2002) to reflect their own state's standards and expectations for student performance (e.g., Florida's Sunshine State Standards).

Research has shown that successful performance on the DIBELS measures provides a reliably predictive method of ensuring students will be successful on later high stakes testing (Barger, 2003; Buck & Torgesen, 2003; Good, et al., 2001; Good et al., 2002; Shaw & Shaw, 2002). Good et al, (2001) found DIBELS Oral Reading Fluency (ORF) measure is a reliable predictor of performance on Oregon's statewide achievement test. Additionally, using the benchmarks of Above 110; 80-109; and Below 80 the researchers correctly classified 96% of low-risk students of failing the statewide achievement test and correctly classify 72% of high-risk students for failing the test. Where students are identified as in need of strategic and intensive interventions additional assessments may be necessary for intervention design in order to best serve the unique needs of the individual student to ensure reaching their benchmark goals.

Shaw and Shaw (2002) evaluated the relationship between the DIBELS ORF measure and the Colorado State Assessment Program (CSAP). Third grade ORF measures for 52 students were analyzed in relation to their third grade performances on the end of the year CSAP. The performance level categories on the CSAP are identified as (1) unsatisfactory, (2) partially proficient, (3) proficient, and (4) advanced. Results indicated the spring ORF measures correlated with the CSAP at .80 indicating a high

relationship with each other. Both the fall and winter third grade ORF measures correlated with the CSAP at .73. Student performances were grouped by CSAP performance levels and analyzed against performance on DIBELS ORF. Of the 30 students who scored at or above 110 on the third grade spring benchmark, 90% performed at or above the proficient level on the CSAP. Of the remaining 28 students who performed below the DIBELS ORF spring criteria of 110 words per minute, only 16 scored in the proficient level of the CSAP. The authors concluded the criterion of 110 words per minute is a sufficient indicator for success on the CSAP. Results also suggested students who performed between 90 and 110 have a high probability of performing at proficient levels on the CSAP because those students who performed above 90 (including those scoring above 110), 91% scored at the proficiency level or above. Only 27% of those who scored less than 90 on the ORF measure performed at the proficient level on the CSAP. The authors argue that an appropriate cut off score for ensuring proficient performance on the CSAP may be as low as 90, however they point out that all of the students who performed at the advanced level of the CSAP has ORF scores of 120 or above.

Barger (2003) compared the DIBELS ORF measure with the North Carolina end of grade reading assessment. The study involved 38 third-grade students who participated from one school and were given the Spring ORF in first week of May. One week later, the same students were given the North Carolina End of Grade Test. This test consists of 56 questions and students have 115 minutes to complete the test. Students read each passage and answer a series of multiple choice questions before moving on the

next passage. The North Carolina End of Grade Test uses a four-level grading scale: Level 1 = lowest level/insufficient mastery; Level 2 = inconsistent mastery; Level 3 = consistent mastery; Level 4 = highest level/superior mastery. Students must achieve a level 3 to be considered on grade level. The author found a cut off score of 100 correct words per minute on the ORF measure was a reliable indicator of obtaining at least a level 3 score on the NCEGT. Correlation between NCEGT and DIBELS was .73 which was similar to the correlations found by Good, et al. (2001) on the Oregon Statewide Assessment ( $r = .67$ ), and by Shaw & Shaw (2002) on the Colorado State Assessment Program ( $r = .80$ ).

Buck and Torgesen (2003) found similar results when DIBELS was used to predict later performance on the Florida Comprehensive Assessment Reading Test ( $r = .73$ ). Buck and Torgesen studied the relationship between the DIBELS ORF measure and the Florida Comprehensive Assessment Test (FCAT). They also examined the relation of ORF measure to the FCAT-math section, and FCAT reading comprehension section, as well as the FCAT-norm referenced test. Scores on the FCAT range from 1 to 5 on with a score of three or higher indicating at or above grade level. Thirteen schools from one school district participated with a total population of students at 1102. Significant correlations were found between ORF scores and the reading FCAT scores ( $r = .70$ ,  $p < .001$ ) and the math FCAT scores ( $r = .53$ ,  $p < .001$ ), and reading scores on the norm referenced test FCAT ( $r = .74$ ,  $p < .001$ ).

The Buck & Torgesen results were consistent to Good et al., (2002) that ORF scores above 110 are considered to be at low risk for reading skills below grade level on a

comprehensive measures of reading comprehension. ORF scores between 80-109 per minute are considered to be at some risk, while scores below 80 are considered at high risk of scoring below a level 3 on the FCAT. Similar to Good et al. (2002), this study found 91% of students who performed at or above 110 on the DIBELS ORF achieved adequate performance on the reading section of FCAT. Eighty-one percent of students who performed at or below an 80 on the DIBELS ORF performed below a level 3 on the reading section of the FCAT. The authors found scores above 110 for minority students was less predictive of success, while scores below 80 for this population more predictive of failure. However, the authors caution the reader about this result considering the small number of minority students involved in study. Authors tentatively suggest the difference in race predictions may be linked to vocabulary development.

#### *DIBELS Best Practice*

Researchers have identified best practices in the use of DIBELS, and they describe its use within an outcomes-driven model (Good, Gruba, & Kaminski, 2001; Kaminski et al., 2008). The primary message provided by these researchers is that schools can prevent reading difficulties for a significant number of students in the elementary school years if schools adopt a prevention-orientated assessment and intervention approach to reading development and programming. DIBELS were developed to provide schools with an assessment tool that is predictive, cost efficient, time efficient, leads to intervention, provides both formative and summative evaluation of student progress, and provides information towards the evaluation of curriculum and instructional decision making (Kaminski, et al., 2008).

The Outcomes-Driven Model is derived from a problem-solving model and the initial application of the problem-solving model to early literacy skills (i.e., Kaminski & Good, 1998). The Outcomes-Driven Model is essentially a decision-making system that focuses on preventative approaches in order to minimize reading difficulties for students. It is also a model that is highly consistent with a Response-to-Intervention (RtI) approach to educational service delivery advocated through the passage of the Individuals with Disabilities Education Improvement Act of 2004 (Kaminski, et al., 2008). Five educational decisions are utilized to accomplish steps to outcomes. Those decisions are: (1) identify need for support, (2) validate need for support, (3) plan for support, (4) evaluate and modify support, and (5) review outcomes.

The basic premise behind the Outcomes-Driven Model is that no child is allowed to fail. The instructional support provided to a student must match their instructional needs. Benchmark testing is encouraged to first identify those students who may be in need of further testing and/or critical instructional support. Students who are identified as at-risk for reading difficulties during the first step of the model may be more closely evaluated to validate their status. Evaluation of human error in the scoring or administration of the DIBELS is primary and should always be the initial assumption during the second step of an outcomes model. When administration and scoring errors are ruled out and/or corrected, students are retested with alternative forms of the DIBELS measures. Students participating in this step of the model are tested at least three times with alternate forms to increase the validity of the scores in determining whether a low score is the result of low skills rather than a “bad day.”

Three patterns of assessment results are possible on repeated assessments: (1) consistently low on all measures, (2) increasing performance with each testing, and (3) extreme variability in results. Where extreme variability is found, other factors potentially affecting the child's performance are evaluated. Students whose need for critical instructional support is validated the third step in the outcomes model is to plan for instructional support. Such a plan should include (1) a clear instructional goal that will be a step to outcomes for the achievement of reading; (2) a focus on essential skills; (3) a plan for the amount and type of support a student needs; (4) a specification of the logistics of who will teach using what instructional materials, when and where; and (5) a measurement plan to evaluate progress. Identification of the goal in relation to baseline measures validated for a student with at-risk status allows educators to visually track or monitor the progress of the student's performance on a graph.

It is important to identify the specific skill(s) that will be directly and systematically taught. It is also important to distinguish between what skill is to be taught, how often instruction is provided, what curriculum materials are to be used, and what instructional procedures are followed. The authors emphasize that a match between student need and instructional support level is critical at this step of the Outcomes-Driven Model. Once a student is provided instructional support, it is then necessary to monitor the progress of the student's performance using the DIBELS measures as a means of evaluating student characteristics of performance and instructional effectiveness. This step is considered key in the Outcomes-Driven Model because formative assessment of

student progress is essential to the efficient and timely modification of instructional supports.

By tracking the student's progress more frequently, it is possible to determine when instructional modifications are needed to assist the student in meeting their goals. When students with substantial need for instructional support are identified, formative assessment should be as often as weekly. Students will require less frequent monitoring where less need is identified. The authors advocate a rule to follow for determining when to adjust instructional or curricular methods as performance falling below the aim-line on three consecutive data points. The purpose of the last step in the model is the review of student outcomes in terms of goal attainment.

Student outcome data may be used to review the structure of supports the school has in place to achieve outcomes at both an individual level and a systems level. At the individual level, a student's performance on the DIBELS measures in relation to their intended goal facilitates the outcomes review. At a systems level, aggregate student data facilitates the evaluation of a school's structure of instructional supports and the overall effectiveness of curriculum and instruction provided in supporting all children to achieve reading outcomes. The authors provide a general rule of 80% as the goal for a school's core curriculum and instruction. The core curriculum and instructional methods should allow for 80% of the student population to meet their grade level goals on the DIBELS measures. It is assumed that at least 15% will require additional targeted support for areas of specific difficulty with another 5% requiring intensive, carefully designed instruction to achieve benchmark goals. The authors describe two reports, the outcomes

report and the benchmark linkage report, which may be used to review a school's reading program or core instruction and additional instructional supports.

An outcomes report assists a district in determining what percentage of students achieved essential reading outcomes as defined by the school district. By having a clear set of achievement goals, a district may review from year to year the progress towards reaching that goal. One method in which DIBELS is designed to assist in this approach is to review aggregated DIBELS data by grade level for each of the measures including Oral Reading Fluency which begins in the first grade and then comparing the percentage of students performing at the bottom quartile, the top quartile, and the distribution of scores in between. Because the DIBELS measures are based in rate of performance, the x-axis of an outcomes graph would display the number of correct words or correct sounds, etc. The y-axis would indicate the frequency of students performing at x-axis levels. The authors indicate that though the outcomes report helps a district stay focused on the "bottom line" it provides little assistance in determining whether the core curriculum and instructional practices are in need of modification.

The benchmark linkage report provides a picture of the linkage between students achieving earlier benchmark goals and achieving later benchmark goals. The goal is to ensure later achievement of benchmark goals by reaching earlier benchmark goals on time. The benchmark linkage report provides what the authors describe as a "dot picture" which shows student performance on an earlier benchmark and a later benchmark at the same time. For example, Figure A4 shows a student's performance on an earlier benchmark along the x-axis (e.g., winter Onset Recognition Fluency) may be compared

to the same students' performance on a later benchmark measure along the y-axis (e.g., spring Phoneme Segmentation Fluency). The graph would also indicate the cutoff scores for each of the two benchmark measures indicating those students who reached benchmark goals and those who did not. The benchmark goal for winter Onset Recognition Fluency is 25 and the benchmark goal for spring PSF is 35. Scores of 10 or less on either measure are considered performances considered as deficit. Scores between 10 and 25 on Onset Recognition Fluency and scores between 10 and 35 on Phoneme Segmentation Fluency are considered performances At-Risk for reading failure.

Four different zones of performance would be identified on the graph. The top right zone, or what the authors identify as Zone A, would be indicative of those students who reached the earlier benchmark goals and the later benchmark goals as well. These students would be identified as those students on track towards successful performance on later high stakes testing. Student performances in the bottom right zone, or Zone B would be indicative of those students who reached earlier benchmark goals but did not achieve later benchmark goals. Students in Zone C, or the top left area of the graph would be indicative of those students who did not meet the earlier benchmark goals, but were able to achieve later benchmark goals, perhaps due to instructional modifications made just following the earlier benchmark testing. Students who did not reach benchmark goals for either of the two benchmark times cannot be predicted to succeed on later high stakes testing and are students whose performances requires intensive remedial efforts. By arranging various combinations of DIBELS measures across different grade levels, a school may more accurately determine the extent to which their core curriculum

and instructional structures are preparing students for later high stakes testing (e.g., third grade FCAT in Florida).

Good et al., (2001) and Kaminski, et al., (2008) suggest a school can identify strengths and weaknesses in their system by reviewing linkages from kindergarten through third grade. When students are identified as performing in a zone of concern (e.g., B, C, or D), a school would first determine if that performance is valid or if testing errors occurred. Where both hypotheses are ruled out, curriculum and instruction factors are considered. According to the authors, examination of the linkage reports allows a review of outcomes in terms of (1) the quality, focus, and intensity of the core curriculum and instruction and the system of providing additional supporting instruction prior to the first benchmark and (2) the quality, focus, and intensity of the core instruction and curriculum and system for providing additional instructional supports between the first and second benchmark.

Finally, Good et al., (2001) and Kaminski et al., (2008) suggest a model for decision-making and assessment. The model contains three primary decision-making loops: (1) Assessment-intervention feedback loop, (2) Review of benchmark goals from one benchmark to the next for individual students, and (3) review of year-to-year outcomes. The first loop involves progress monitoring interventions for those students identified early in the year as experiencing reading difficulties. The second loop involves analysis of individual student performances through the year from one set of benchmark measures to the next. Students receive additional instructional supports in the beginning of the year due to at-risk concerns identified from the first benchmark, but who perform

successfully on the next set of benchmarks would not require the same level of supports and monitoring as long as they continue to achieve adequate progress in subsequent benchmark goals for the year. The third loop describes the process of using outcome reports and benchmark linkage reports in order to evaluate the core curriculum and instruction and system of providing additional instructional supports of a particular school.

Other studies have also advocated for the use of DIBELS as an effective and efficient assessment approach to building and choosing effective curriculum and instructional practices in the school and classroom environments (Simmons, Kame'enui, Good, Harn, Cole, & Braun, 2001; Smith, et al., 2001). For example, Simmons et al. (2001) discussed the utility of DIBELS assessment in building, implementing and sustaining a beginning reading improvement model at a school level. Overall, the authors describe implementing large scale reading programs in a number of schools. To start, the authors advocate three general principles concerning the building, implementing, and sustaining of an effective reading program. First, where schools vary in size and in amount of resources, “the principles and strategies of conceptualizing, designing, implementing, and sustaining instructional and behavioral change are fundamentally the same for all individual schools (p. 539).” Second, if effective reading programs are to be sustained over long periods of time, they must be implemented, monitored, and supported at the school-building level. And third, implementation and support of comprehensive programs at the building level are a necessary, but insufficient condition for increasing

and sustaining student performance. Thus, district level support and commitment are necessary for long-term sustainability.

The authors state the implementation of a beginning reading program involves knowledge of reading in an alphabetic writing system and procedural knowledge of how to organize and implement research-based reading programs in schools which are composed of people, practices, pedagogy, and policy. An additional eight research-based tenets serve as the basis for the design of a school-wide reading program. First, address reading success and failure from a systems level perspective. It is important for school personnel to recognize how their policies and practices contribute to, inform, and/or detract from their efforts to delivery effective reading instruction. Next, embrace a preventative approach by focusing on early assessment and early intervention. Recognize and respond to multiple contexts of reading achievement. Utilize a research-based core curriculum and enhancement programs. Be familiar and utilize the convergent knowledge of effective reading practices. Use school-based teams to customize interventions. Rely on and foster ability of the principal as instructional leader. Finally, use formative assessments that are sensitive to changes in student performance for early identification and evaluation of interventions.

The model articulates five distinct stages for implementing and sustaining a beginning reading program. The first stage consists of auditing a schools' current reading program and to assess baseline student performances. The focus at the school level is the use of the Planning and Evaluation Tool (Kame'enui & Simmons, 2000) to conduct the

school audit. The student level focus is on the use of DIBELS and CBM for determining baseline student performances.

During Stage 2, the school level team reviews the audit, identifies strengths and areas for development, prioritizes development plans, and establishes an action plan. At the student level, individual students are identified as either meeting benchmark goals, at-risk for reading failure in later years unless provided with strategic intervention planning, or in need of intensive interventions to remediate deficit levels in performance.

During Stage 3, the building team designs the core instructional interventions by specifying the goals of the school, core curriculum program to be used, time allocated for reading, instructional grouping and scheduling, instructional implementation and training, and design of a progress monitoring system. At the individual level, intervention planning is conducted for those students requiring either strategic or intensive instructional supports to meet later benchmark goals. The teacher or team specifies individual goals, the curricular and instructional methods to be used for intervention, the amount of instructional time afforded, and a system for monitoring student performance.

Stage 4 involves implementing a progress monitoring system towards stated goals at both the school level and the individual student level. Additional focus at the school level concerns identifying valid and reliable indicators for evaluating program effectiveness, commit resources, determine schedule for monitoring, and interpret and communicate results. At the individual level students identified as requiring intensive interventions are in need of weekly to bi-weekly monitoring, whereas students identified

as merely at-risk may only require monthly monitoring. Benchmark testing resumes for all each of the four times scheduled throughout the school year.

The final stage involves an evaluation of achievement towards goals at both the school and student levels. At the school level, the authors suggest engaging in this stage at least three times during the year in respect to both intermediate goals and end of the year goals. Additional analyses may involve an examination of intervention components at the building level, adjustments of instructional practices and curricular materials, as well as a determination to continue key interventions. At the student level individual students are evaluated in terms of achievement relative to their goals. Students not learning enough should have instructional profiles charted to determine amount of growth achieved and to determine whether there is a need for reduction, continuation, or the intensifying of intervention strategies. DIBELS and CBM-Reading are utilized through this whole process at each stage of the model at both the school and individual levels. Aggregated DIBELS data for the school is utilized to make decisions at that level while individual student data are utilized to monitor progress towards each of the scheduled benchmarks through the year. The authors emphasize that use of these measures does not negate the use of other measures or reading dimensions such as vocabulary and comprehension. Instead, student performance on these two measures is used as indicators for efficient and reliable progress monitoring, and outcomes assessment.

Similarly, Smith, et al. (2001) supported the use of DIBELS has having high treatment validity. These researchers found two effective dimensions of a professional development program in early literacy skills that lead to significant increases in student

reading achievement performances: (1) adoption of a dynamic early literacy assessment system, and (2) teacher implementation of research-based early literacy practices. Their article focused on a case study involving an evaluation of the link between teachers' conceptual knowledge about reading development and instruction and student performance on DIBELS. Eighty-five percent of the students who participated in the study were identified as having free/reduced lunch status, which was the highest percentage in the school district. Approximately 15-20% of students were English-language learners. The summary of the article discusses the results in student achievement after 4 years of continued professional development at the school.

The authors found it is the way the curriculum is implemented rather than the type of curriculum per se that seems to make the difference according to the teachers at this school. Authors argue that changes in teacher's knowledge of research-based practices in teaching reading led to improvements in instruction, which when combined with an ongoing formative assessment of student performance led to the significant increases in student reading achievement scores. The 4-year professional development program used a series of professional development meetings to gradually introduce teachers to relevant aspects of the early literacy research base. Teachers were asked to prepare for each meeting by reading a selected research article or a one-page synthesis the authors prepared of a report or article. Meetings also provided teachers with opportunities to have discussions about student responses to literacy instruction, determining which practices to keep, modify, or discontinue, and how to align new and existing knowledge about the teaching of literacy skills in kindergarten.

A variety of themes were discussed during the trainings including big ideas in early literacy, explicit and consistent teacher wording, sequence of letter name and sound instruction for early use of words (e.g., use of high utility letters and first), emphasis on systematic review for mastery of skills, establishing appropriate goals and using ongoing assessment to inform decision making, and implementing regular testing. The authors also listed 7 critical components of early literacy instruction.

1. Allocated time for daily highly focused literacy instruction
2. Consistent routines for teaching big ideas for early literacy
3. Explicit instruction for new letter names and sounds.
4. Daily “scaffolding” or assisted practice with auditory phoneme detection, segmentation, and blending.
5. Immediate corrective feedback.
6. Daily application of new knowledge at the phoneme and letter-sound levels across multiple and varied literacy contexts.
7. Daily review

Good, et al., (2003) described how to use DIBELS to evaluate a schools’ core curriculum and system of additional intervention in kindergarten. The authors indicate the article draws heavily from the information available from schools and districts participating in the DIBELS Data System; a longitudinal data collection system of aggregated school data and student performances using DIBELS. The DIBELS Data System at the time of the article was comprised of 300 school districts, more than 600 schools, and more than 32,000 students during the 2001-2002 school year. The authors

caution the reader that (1) the data on the DIBELS Data System may not be representative of the nation at large, and (2) there are currently no procedures in place at participating schools to ensure adherence to standardization practices in the use of DIBELS, thus the data obtained from schools must be analyzed cautiously.

The authors provide a template for formatting school reports. The authors identify important participants in the DIBELS system as school principals, kindergarten and first grade general education teachers, remedial teachers, school psychologists, speech-language pathologists, and other support personnel as available in the school. The authors also identify a number of critical components to an effective beginning reading program. Consistent with Smith, et al. (2001), the authors advocate for a linkage between professional development, realities of the classroom, and the use of student data to give new teachers ways of reflecting on their teaching and choice of materials. School-based reports are encouraged for use in operationalizing four principles of effective professional development: (1) identifying appropriate benchmark goals; (2) links between technical and conceptual components of instruction, early literacy skills, and DIBELS measures; (3) grade level discussion support systems to enhance instructional programming; and (4) use of visual representations to assist teachers in monitoring the effectiveness of their curricular and instructional choices in relation to student performance.

DIBELS may be used to make decisions about core instructional practices and curricular materials. The authors argue that to meet kindergarten benchmark goals, a number of supplemental materials and programs are necessary. The authors advocate for research-based programs which offer the general education kindergarten teacher a variety

of options to support the core lessons of the classroom. School reports are also encouraged for use in evaluating kindergarten curriculum in determining the effectiveness of the core curriculum and instruction. The authors caution that a supplemental reading programs' evidence of effectiveness does not guarantee its successful use with every child needing intensive support. Frequently monitoring the progress of such student helps to ensure the appropriate type and amount of instruction and curriculum supports are provided.

Consistent with Good, et al. (2001), the authors describe the use of DIBELS in an Outcomes-driven Model. Extending this work, the authors of this chapter provide 7 organizing questions for schools when analyzing student data for their school. For example, how effective is the core curriculum and instruction in supporting students who are entering kindergarten with benchmark skills to achieve the DIBELS ISF goal in the middle of kindergarten? How effective is the system of additional interventions in supporting students who are entering kindergarten at risk for reading difficulty to achieve the DIBELS ISF goal in the middle of kindergarten?

If schools are interested in factors related to the community context, effectiveness of preschools in the community, and how much emphasis on early literacy skills exists in the community, then schools could compare their entering kindergarten students' DIBELS results with those from other schools. The entry-level skills can also reflect the language and cultural factors within the larger community. Schools may use early kindergarten DIBELS data (ISF and LNF scores) to assess the degree to which students require intensive instructional supports. Where significant numbers of students (approx.

more than 20 percent) require additional or intensive supports, a school should review and adjust its core curriculum and system of additional interventions. The authors provide a table for schools to compare their kindergarten students' entry level achievement performances with other schools participating in the DIBELS Data System; a nationwide collection of DIBELS data.

With regards to mid-year performances of kindergarten students, the authors suggest a student with risk indicators in two or more areas may require intensive interventions to achieve early literacy goals. It is also important to analyze mid year patterns of student performance on the DIBELS measures to identify any measurement error or problems with the integrity of the assessment process. Three types of analyses may be utilized within the context of answering this second question. First, how well do the current kindergarten students perform mid-year in relation to previous years? Second, how well do the current kindergarten students perform mid-year in relation to other schools in the DIBELS data system? Finally, how well do current kindergarten students perform mid-year in relation to the schools desired kindergarten mid-year goals?

If a school wishes to examine the effectiveness of the core curriculum and instruction in assisting most of the students in reaching the mid year benchmark goals in kindergarten, DIBELS data may be used to facilitate such an examination. An effective and appropriate core curriculum and instruction should ensure that the majority of those students who reached beginning of the year benchmarks are able to also reach mid year benchmarks and end-of-the-year benchmarks. The authors provide a table for normative assessment using other schools in the DIBELS Data System. The authors advise that

given the normative data available in the DIBELS Data System, there exists marked variability in schools' core curriculum and instruction in supporting those students who met beginning of year benchmarks to achieve mid year benchmarks. The authors suggest it appears most schools do not *adequately* address the skills of initial sounds in the first half of the kindergarten year.

DIBELS data can also be used to examine how effective the system of additional interventions is in supporting students entering kindergarten who are at risk to meet mid year benchmark goals. The normative data provided by the authors is alarming in that a typical school in the DIBELS Data System supports only 3% of the students with an intensive instructional recommendation at the beginning of the school year for kindergarten to meet mid year benchmark goals. The normative data also suggest 45% of schools do not support any of the students with a recommendation at the beginning of the year for intensive instructional support to meet mid year benchmarks. This does not necessarily mean that school are not doing anything to help these students, but that rather many of them are not benefiting from what services are provided to meet mid-year benchmarks. In the DIBELS Data System, focus for end of year kindergarten goals has been on PSF and NWF measures, while treating LNF as a risk indicator. The authors state while the research shows naming letters is a good predictor of later literacy skills, why it is a good predictor is less clear. Benchmark achievement at end of kindergarten year seems to be a good predictor of end of year first grade performance. The DIBELS Data System demonstrated that 87% of students who reached end of kindergarten goals reached the DIBELS ORF goal of 40 words per minute or more at end of first grade.

The DIBELS Data System may also be used to address how well a school compares to other schools in the DIBELS Data System on end of year kindergarten goals. In a typical median school, 15% of students have a recommendation for intensive supports, 17% are identified as requiring strategic supports, and 65% are identified as requiring benchmark general education core curriculum and instruction. The authors provide a normative table for schools to evaluate their end of kindergarten outcomes compared to other schools. The sixth organizing question addresses the effectiveness of the core curriculum and instruction in supporting students who are met mid year benchmarks to achieve DIBELS PSF goal at the end of kindergarten year. The primary purpose is to identify areas for improvement in the kindergarten program. The authors suggest disappointing end of year outcomes may be attributable to (1) low early literacy skills in the middle of kindergarten, (2) inadequate focus provided by the core curriculum and instruction to emphasize essential components of early literacy skills, or (3) the system of additional interventions is not providing adequate support for students identified at-risk.

Finally, the DIBELS Data System can offer a school an opportunity to examine a school's system of additional interventions in supporting students identified at-risk for middle of year kindergarten goals to achieve end of kindergarten PSF goals. Normative data is again provided for schools to evaluate their system of supports in comparison to other schools. The normative data suggests typical schools in the Data System can support approximately 26% of these students. Many schools however are able to support at least 91% of students with intensive needs mid year in kindergarten.

The authors close by providing a few observations from their experience in helping schools utilize DIBELS to improve schools' abilities in meeting their literacy goals. First, data are helpful because teacher perception is not always accurate. Second, changes in outcomes at one grade level precipitate changes in the next grade level. Next, grade level data across classrooms indicate much about the general way of doing business within a school. Also, outcomes are stable and replicable unless big changes in curriculum, instruction, and system of additional intervention are made. Finally, even when the schools have very different orientations to beginning reading instruction, evaluation of student outcome data can be used by schools to change reading outcomes.

#### *Teacher Training and Data Utilization*

In addition the general findings provided by Chen and Rossi, (1983) regarding implementation of programs in an educational setting and results of the study conducted by Smith, et al. (2001), additional research has investigated the link between teacher utilization of assessment data to make instructional decisions and the degree of training necessary to improve student outcomes (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1982; Fuchs, Deno & Mirkin, 1984; King, et al., 1983; Wesson, et al., 1983). The combined results of these studies suggest the provision of assessment data alone is insufficient for ensuring teacher utilization of data to make instructional changes to improve student performance. Furthermore, teachers require high levels of support and ongoing professional development in addition to technically adequate assessment data to ensure high quality instructional changes are made to improve student performance.

Fuchs, et al. (1982) investigated how the introduction and use of data-utilization strategies affect the number of modifications teachers make in their classroom and student performance. Teachers were trained during a one week schedule of full day workshops prior to the start of the school year and during bi-monthly, half-day workshops throughout the school year. The focus of training was on training teachers to write curriculum-based IEPs, create a curriculum-based measurement procedure, measure frequently and graph student progress toward IEP goals, and develop strategies to improve the feasibility of implementing the frequent measurement systems. Result indicated that measurement of student progress alone was insufficient for ensuring that those data will be used to make instructional changes in students' educational programs. Authors concluded teachers will require specific evaluation procedures and data-utilization rules to ensure that teachers will use data to make instructional changes to increase student performance.

In a study investigating the effects of repeated and frequent testing on student academic achievement, Fuchs, et al. (1984) found repeated use of CBM-Reading combined with evaluation procedures for data analysis affected positively both student achievement and student awareness of their own achievement. Students in the experimental group performed better than student of comparison teachers on virtually all achievement measures, rate and accuracy of oral reading scores, and performance on the *Standford Diagnostic Reading Test*. What is most interesting about this study is that teachers were not instructed on what changes to make in response to assessment data. Teachers were merely scheduled to evaluate data trends and levels at frequent intervals

and “introduce a new dimension” into the student’s program where performance was found to be unsatisfactory.

### *Response to Intervention (RtI)*

Passage of the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA 2004) in 2004 launched a movement across the country for incorporating a new model of service delivery and determination of special education eligibility. At the heart of this new initiative is the use of a problem-solving model for achieving better educational outcomes for children (Castillo, Cohen, & Curtis, 2006). Decisions made towards eligibility for students suspected of having a learning disability would require the evaluation of a student’s response to interventions provided in the school. As Castillo et al., indicate, this form of service delivery involves substantial systems’ change. This national movement of systems change is expected to incorporate evidence-based instructional practices for all students and the use of progress monitoring data to evaluate student growth in academic outcomes such as reading.

In a review of four different types of approaches to the assessment of children and adolescents with learning disabilities, Fletcher, Francis, Morris, and Lyon (2005) reported that a Response-to-Intervention model was more reliable and valid than the use of traditional aptitude-achievement discrepancy, low achievement, and intra-individual difference models. However these researchers advocate for a combined RtI and low achievement model for assessing students suspected of having learning disabilities. The general concept of an RtI model of is an approach to linking several short assessment probes while intervening with a child in a specific content area (Fletcher et al., 2005).

Thus, underachievement may then be operationalized, in part, by the relative responsiveness of the student to empirically-validated interventions provided to the student (Gresham, 2002).

Griffiths, VanDerHeyden, Parson, and Burns (2006) discussed the practical applications of a RtI model. They characterize an effective RtI model as having three specific features: (1) systematic data collection to identify students who may be at risk for learning difficulties; (2) effective implementation of interventions for adequate durations, and (3) a review of student progress data to determine level of treatment and density of services. These researchers further identified two basic approaches to RtI in their review of the relevant research literature: (1) standard protocol; and (2) problem-solving model. At the time of the present research study, a statewide program evaluation study was underway towards an identification of critical elements needed for successful implementation of a problem-solving RtI model of service delivery in Florida (Batsche, Curtis, Dorman, Castillo, & Porter, in press; Castillo, et al., 2006). In short, a problem-solving/RtI model of educational service delivery is very consistent with Kaminski, et al.'s (2008) Outcome-driven Model for using DIBELS data and holds promising potential for educators towards increasing student achievement outcomes and treating academic difficulties through a prevention-oriented approach.

## CHAPTER THREE - METHOD

### *Purpose*

The present study was conducted to provide information that may be useful for improving implementation on the use of DIBELS, features regarding DIBELS data collection and analysis methods at the school building level, and Reading Coach consultation efforts. The results were aggregated and summarized in a manner that allowed for the understanding of what was occurring in the field through a description of interviews and focus groups, as well as identifying salient issues or variables for future research. Overall, the research literature suggests teachers' skills in the use and analysis of data to guide instruction is a topic that must be understood (Kerr et al., 2006; Smith, et al. 2001). Therefore, the general purpose of the present study was to understand teachers' perceptions and uses of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) within a *Reading First* program.

### *Research Questions*

Based on the research literature and the purpose of the study, three research questions were formulated related to teachers' use of DIBELS data:

1. What are teachers' perceptions and understanding about DIBELS and the PMRN?
2. How do teachers' understanding and use of DIBELS data, as presented in the PMRN reports, compare to *Reading First* experts who are provided with the same information?

3. What attitudes and perceptions exist among persons other than teachers who participate in the collection, input, and analysis of DIBELS data throughout the school year?

### *Research Design*

The following study used a basic interpretive and descriptive design (Merriam, 2002a; 2002b) for the purposes of understanding teachers' perceptions and use of DIBELS. Data were collected through a combination of semi-structured interviews, case study interview/observations, and focus groups. Throughout the data collection process contrary evidence was explicitly probed depending on participant responses to interview/focus group questions. Data analyses involved several procedures which may be classified as early concurrent analyses during data collection efforts and later descriptive analyses. Coding of data involved first level coding and pattern coding procedures (Miles & Huberman, 1994) combined with constant comparison methods (Patton, 2002). Within-case and cross-case comparisons were conducted in order to identify convergence of themes and observe variability among responses. A number of research strategies were used to ensure validity and reliability including inter-rater agreement checks, triangulation across multiple sources, member checks (e.g., participant feedback on accuracy and completeness of comments received), description of researcher's position, peer review/examination, data saturation, and an audit trail. The units of analyses in the present study were both the individual and group level. Individual teachers participated in a one-time semi-structured interview and case study review and results were analyzed at both the individual and group level. Reading

Coaches and specialists participated, respectively, in a one-time focus group and results were analyzed at the group level. Experts participated in a one-time case study review and results were analyzed at the individual and group level.

### *Participants and Sites*

#### *Description of Sites*

Information available through the Florida Department of Education (FLDOE) identified the participating school district as one of the largest school districts in the State of Florida, serving a student population larger than 100,000 students. The school district employed approximately 14,500 employees across 169 schools. At the start of the present study, 54 elementary schools, of the 87 total elementary schools, in the participating district were identified as *Reading First* schools that were in their 4<sup>th</sup> year of implementing the *Reading First* grant. Participant views and comments were sampled across approximately 28% of these schools (i.e., 15 elementary schools). A total of 14 teachers were sampled across 9 elementary schools. Eight Reading Coaches were sampled across eight elementary schools. Six “specialists,” due to their itinerant status, held positions across 13 different schools but indicated only participating in either the input or analysis of DIBELS data at seven schools, collectively. Table 3 shows the representation among the participating schools of teachers, Reading Coaches, and specialists.

#### *Description of Participants*

The participants for this study were seven kindergarten and seven first grade teachers, eight elementary school site-based Reading Coaches, two experts on the use of

DIBELS and the Progress Monitoring and Reporting Network (PMRN), and six school specialists (e.g., non-teaching support service staff who were involved in the collection, input, and analysis of DIBELS data at *Reading First* schools). “Experts” in the present study were selected based on involvement in statewide efforts to provide training and technical assistance to school districts in the use of DIBELS and the PMRN.

“Specialists” in the present study, all combined, represented the roles of school psychologist, academic diagnostician, and ESOL instructional support (an instructional support program for students for whom English is a second language). These individuals held itinerant positions across several schools; *Reading First* and non-*Reading First* schools.

Demographics were collected for all teachers who participated in the present study. All teachers were female and the reported years of experience ranged from 2-26 years at their present grade level at their present school. Half the teachers (seven) reported less than 10 years total experience teaching reading in elementary schools while the remaining seven reported experience teaching reading in elementary schools above 10 years. Age of teachers ranged from 21-65 years of age. Teachers identified themselves as Caucasian (13) and African American (1). Range of educational credentials included Bachelor’s degrees in K-12 Education (11) and Master’s degrees in education (3).

#### *Sampling Methods and Rationales Used for Selecting Participants and Settings*

##### *Selecting the School District*

A combined purposeful criterion sampling and purposeful convenience sampling method (Patton, 1990) was used to identify one school district which had implemented

the *Reading First* program for at least two years. The use of the convenience sampling method was intended to keep resource costs down associated with travel and time required for conducting the study. The selection of a school district using a convenience sampling method was based upon distance and approval for conducting the present study. The criterion of a two year minimum implementation was established based on an assumption that an initial amount of time was required for a school district to align its resources and standardize its implementation efforts district-wide before any evaluation of the program or its components may be appropriately conducted.

#### *Selecting School Sites*

A purposeful criterion sampling method was used to select potential sites for recruiting teachers, Reading Coaches, and specialists. First, all *Reading First* schools for the chosen school district were identified. Of those schools, 54 were identified as having implemented the *Reading First* grant for more than 2 years.

#### *Selecting Teachers, Reading Coaches, and Specialists*

A purposeful criterion sampling method combined with participant self-selection was used to recruit teachers, Reading Coaches, experts, and specialists. Teachers, Reading Coaches, and specialists were invited to participate in the study by providing invitations to all 54 *Reading First* schools identified as implementing the *Reading First* grant for more than 2 years. The criteria for participation were a minimum 2 years experience working at one of the 54 *Reading First* schools and 2 years minimum experience teaching at current grade level in either kindergarten or first grade (teachers only). These two criteria were established because it was necessary to ensure that

participants had sufficient time to participate in their school's *Reading First* implementation efforts. It was the primary investigator's judgment that two years experience would be a sufficient minimum criterion. Specialists also had to have been involved in either the data collection and/or analysis of DIBELS data at their assigned school.

Kindergarten and first grade teachers were targeted for participation because the combined DIBELS measures used at those grade levels encompass the entire DIBELS assessment. When used as a benchmark screening tool, second grade only uses the Nonsense Word Fluency and Oral Reading Fluency subtests; third grade and above primarily uses the Oral Reading Fluency measure for benchmark screening.

#### *Sample Size*

*Teachers.* The number of teacher participants selected for the present study was determined by using a combined incidence probability sampling method (DePaulo, 2000) and a saturation method (Fossey et al., 2002; Patton, 1990; 2002). Appendix B2 shows a probability sampling table developed by DePaulo (2000) which may be used to determine a minimum starting point for selecting the number of participants for a given qualitative study. Use of this probability table required an initial assumption about the incidence rate of a dissenting or favorable point of view. Informal discussions held by the researcher with Reading Coaches and school psychologists who work in *Reading First* schools revealed, anecdotally, that approximately one-fourth to one-half of the teachers in *Reading First* schools they worked articulated positive perceptions about the use of

DIBELS assessment. This suggested an incidence rate of at least 25% favorable opinion regarding the use of DIBELS.

Based on this assumption, and by use of DePaulo's probability table for qualitative sampling, selecting a random sample of at least 10 teachers should allow for collection of all relevant, and potentially distinct, perceptions with only a 5% probability of not identifying a unique perception for inclusion in the study. Further dialogue with Reading Coaches in the field prior to conducting the study suggested the knowledge needed to understand and use DIBELS among kindergarten and first grade teachers was not unique to their grade level. Thus, teachers were viewed as one group rather than two separate groups.

A saturation method (Fossey et al., 2001; Patton, 1990; 2002) was utilized beyond this minimum number to ensure no new information was available and to add credibility to the selection process. The following set of procedures was followed for determining the need for additional participants beyond the established minimum. First, following collection and an initial review of the first 12 teacher interviews and case study observations (six kindergarten and six first grade), two additional teachers were recruited. If information obtained from either of these two teachers yielded any new information or perspectives, then an additional two teachers would have been selected. These steps would have continued until a saturation point had been reached. A total of 14 teachers (seven for each grade level) were sampled in accordance with the above sampling size methods used as no new information was obtained from the last two participating teachers.

*Experts.* Because teacher perceptions were the primary focus of the study, and because experts were only needed to provide an expert opinion standard for comparing teachers' perceptions and comments regarding the case study presented to them, it was determined that it was not necessary to sample several experts for comparison with teachers. However, it was decided that at least two experts should be sampled in the event that a given expert may make recommendations or comments unique to themselves.

*Focus Groups.* A minimum of five persons and a maximum of ten persons were used as the criteria for conducting the two focus groups (Patton, 2002). This sample size was selected to ensure a minimum level of diversity in views was obtained while also avoiding the potential of having a group too large that it would adversely affect the quality of responses among participants. Additionally, because teacher perceptions and use of DIBELS was the primary focus of the present study, the use of other non-teaching participants was to provide alternative perceptions with which to compare teachers' responses using a triangulation method using multiple sources (Patton, 2002).

#### *Data Collection Instruments*

##### *Semi-Structured Interview Guide – Teachers*

Appendices C through F contain the instruments used in the study. Appendix C illustrates the original interview guide developed for use during the semi-structured interviews with teachers. To answer the first research question, the focus of the interview was to address (1) topics identified as important for understanding based on a review of the relevant research including specific knowledge about DIBELS, level of training on the utilization of DIBELS data, and the climate or culture of supports in relation to

DIBELS and assessment in general; and (2) any unique perceptions or use of DIBELS emerging from interviews with teachers. Changes or emphases made to the interview guide in response to teacher comments were analyzed and are included in the Results section of this report. The Teacher Interview Guide was intended to be approximately thirty minutes in length. Only one version of this form was used as the questions asked of teacher participants were not exclusive to any particular grade level teacher.

#### *Case Study - Teachers and Experts*

The case study materials were developed to answer the second research question. Appendix D illustrates the Case Study Questions and PMRN data reports used in the present study for conducting the teacher and expert case study review. Follow-up questions listed were used as needed based on teacher/expert comments. The case study review involved a kindergarten and first grade version depicting a particular student at the end of the year (i.e., 4<sup>th</sup> assessment cycle). Teachers were given the case study corresponding to the grade level they were teaching at the time of the interview. Three PMRN reports involving real student data were utilized to share with teacher participants. All identifying information for the case study student, class, school, and district were removed for confidentiality purposes. The case studies involved the following specific reports for use: Class Status Report (Appendix D2 & D5), Student Grade Summary Report (Appendix D3 & D6), and the Reading Progress Monitoring Student Cumulative Report (Appendix D4 & D7). These reports were chosen from the variety of reports available through the PMRN based on recommendations by Reading Coaches currently

serving *Reading First* schools and who work with teachers on a daily basis. The case study review was intended to take approximately thirty minutes to complete.

*Description of PMRN Reports.* The Class Status Report showed how well students in the class performed on a current DIBELS assessment cycle. This report allowed the user to organize the class list based on either alphabetical order or in order of “Recommended Instructional Level” (i.e., intensive, strategic, or initial). For purposes of this study, the class list was organized in order of “Recommended Instructional Level.” The color coding system used on this type of report also communicated the level of concern regarding a student’s performance. RED indicated a student’s performance for a given subtest was well below expectations. YELLOW indicated an at-risk performance. GREEN indicated a student’s performance was considered meeting benchmark expectations while BLUE indicated a student performance that may be considered above average. The same color coding system was used for the “Recommended Instructional Level” indicator except that BLUE was not used. Students demonstrating a profile that was above benchmark expectations for the particular assessment cycle were listed as GREEN “Initial” as the “Recommended Instructional Level.”

The Student Grade Summary Report provided a comprehensive summary of a student’s performance on the most recent assessment as compared to the class. Box Plot graphs were designed to communicate class performance between the 25<sup>th</sup> and 75<sup>th</sup> percentiles, including a horizontal line within the box which communicated the median class score. Whiskers indicated student performances above the 75<sup>th</sup> percentile or below the 25<sup>th</sup> percentile. Individual data points above or below the whiskers represent outliers

whose scores fell within the upper or lower 5<sup>th</sup> percentiles. These box plots were depicted in relation to the benchmark goal for the specific subtests used in the assessment cycle designated by a solid green line. The graph allowed a teacher to evaluate a student's performance in relation to both the class average and the benchmark expectation. The student's position in relation to the class was indicated by a small colored flag with letters indicating the student's risk status: HR for High Risk, MR for Moderate Risk, LR for Low Risk, and AA for Above Average. The same color coding was applied to the student flag.

The third report was the Reading Progress Monitoring Student Cumulative Report. This report provided a set of tables which showed the student's progress on each subtest given throughout the school year for each quarter of benchmark assessments. This report also provided an overall Recommended Instructional Level (RIL) for each assessment period. Instructional level descriptors used were Initial Instruction (ii), Strategic Instruction (S), or Immediate Intensive Instruction (iii). End of year outcome tests such as the Peabody Picture Vocabulary Test (PPVT) and the Stanford Achievement Test – Tenth Edition (SAT-10) were also provided for 3<sup>rd</sup> cycle reports. This report allowed a teacher to evaluate the progress of a student on a particular subtest/skill (e.g., Nonsense Word Fluency) and identify their change in performance status as the year progressed (e.g., strategic instruction).

#### *Focus Group Guide*

Appendix E illustrates the Focus Group guide used for investigating the third research question involving two separate focus groups; one with Reading Coaches and

one with specialists. The questions used were developed to understand non-teacher perceptions and use of DIBELS, as well as non-teacher perceptions of teacher use of DIBELS. Each focus group was intended to be approximately one hour and thirty minutes; the final thirty minutes was used for summarization of findings based on notes taken during the session for validation purposes.

#### *Instrument Validation Procedures*

A peer review/examination method (Merriam, 2002b) was used to field test the above instruments for use prior to data collection. Peers in this context included: Researchers experienced in qualitative research methods; researchers knowledgeable in the use of DIBELS; Reading Coaches and a Reading Coach supervisor; and elementary teachers. The comments and feedback received by all reviewers were not recorded or used as primary data for the present study; though they were used to modify instruments as needed.

A district-wide *Reading First* supervisor provided feedback about the interview guide, case studies and the focus group guide. Recommendations and changes needed as a result of consultation with the *Reading First* supervisor were made to the interview guide, focus group guide, and case studies. The *Reading First* supervisor was asked to review all instruments for clarity, appropriateness of questions, and relevance to the *Reading First* model.

As a priority, comments were solicited on the equivalence between the two case studies. This was important as it directly affected whether the kindergarten and first grade groups were treated as independent groups. The issue of equivalence for this study

concerned the degree to which a set of skills needed to analyze one case study were sufficient for analyzing the other. Feedback received by the *Reading First* supervisor confirmed the equivalence of the two case studies. Therefore, only one case study guide was needed for use by both kindergarten and first grade teachers.

The Interview Guide and Case Study Questions were field-tested additionally by inviting one kindergarten and one first grade teacher to review the questions and provide feedback on the clarity and appropriateness of the questions. Teachers validated the appropriateness and clarity of the questions being asked within the instruments.

The focus group guideline was additionally reviewed by a researcher at the local university who had expertise in conducting focus groups in order to determine the appropriateness of the questions, the structure of the format, and any other critical information that would ensure an appropriate and credible focus group process. Revisions were made by the researcher following the review of the Focus Group Guide by the university researcher.

### *Data Collection and Analysis Procedures*

#### *Site Entry*

Due to the participating school district's IRB review, all recruitment solicitations had to be provided to all potential school sites and approved by school principals before obtaining consent for participation from any potential participant. Further, the researcher was prevented from identifying and contacting any potential participant directly. School district research compliance requirements also required a district employee to sponsor the research study. For this purpose, a school district *Reading First* supervisor was contacted

for purposes of providing support (see above). Upon receipt of support for conducting the research study it was agreed that all results would be shared with the district and participants. Additionally, the district *Reading First* supervisor offered to contact all 54 schools through e-mail to announce the start of the research study and to acknowledge the district's support for the study.

Following the introductory announcement provided by the district *Reading First* supervisor, packets of invitations for each of the three groups of persons identified above, and including building administrators, were sent to all 54 *Reading First* schools identified as eligible for participation. The following materials were included in each packet of information: (1) Principal invitation cover letter with blank consent form addressed to the school principal introducing the research study and its purpose along with solicitation for support in distributing enclosed materials and participation in a principals' focus group, (2) teacher invitation cover letter and blank consent form, (3) Reading Coach invitation cover letter and blank consent form, and (4) specialists cover letter and consent form. With the exception of the Reading Coach, 5 copies of the teacher cover letters/consent forms as well as specialists' cover letters and consent forms were provided for distribution by the school principal. All invitations provided contact information using a personal cell phone and email address for the researcher. All potential participants were encouraged through the invitations to contact the researcher if they were interested in participating or if they had any questions about the study.

### *Obtaining Consent to Participate.*

*General procedures for obtaining consent.* In general, all individuals asked to participate in any aspect of the present study were required to give written consent using a consent form approved by the supervising university's institutional review board (IRB) for the protection of human research participants. The consent form was written in full accordance with the university's IRB policies. Overall, participants were given a general overview of the purpose and focus of the study. All information identifying any participant, including organization and school site name, were kept strictly confidential and known only by the principal investigator. The consent form detailed the procedures that were followed to ensure a high level of confidentiality.

The consent form also provided the participants with contact phone numbers and email addresses for this investigator as well as supervising personnel from the university. A statement of all risks and benefits for participating in the study as well as informed notice of a provision to refuse participation at any time was included in the consent form. Additionally, to ensure confidentiality, specific contact names and contact numbers or email addresses of all participants was secured using a desktop computer only accessible by the principal investigator and which involved use of a password and username for access.

*Obtaining consent from Reading First supervisor.* One *Reading First* supervisor was asked to participate in the review of instruments for quality assurance and participate in member check procedures detailed below. Upon written consent, the *Reading First*

supervisor was asked to comment on the instruments as described above and later asked to engage in member check procedures to provide validation of findings.

*Obtaining consent from teacher participants.* Each teacher who responded either by email or by phone and who expressed an interest to participate was contacted to schedule a date and time that was convenient for them to participate. No travel was required by any teachers for purpose of this study. At the start of each individual interview with a teacher, time was taken to go through the consent form and provide full disclosure of the study's title and purpose. However, teachers were not informed at first that experts would be involved in the study and asked to provide comments towards the case studies for the purpose of comparison with teacher responses. Following all data collection, teachers were then informed of this procedure and all results shared with them (see below). Teachers were asked to provide written consent on the day of the scheduled interview if they chose to participate.

*Obtaining consent from DIBELS experts.* Contact was made with a representative of a state agency providing training and technical assistance to schools on the use of DIBELS and PMRN. A description of the present study was provided and a copy of the appropriate consent form was faxed to the agency. Permission was provided to contact two experts who had sufficient expertise and knowledge of DIBELS and PMRN. Each expert was contacted by phone separately and asked to participate in the present study. All questions were answered and full disclosure of the study's purpose and scope were given. Upon verbal consent, a copy of all materials associated with the case study review, including a copy of the consent form, was mailed to each expert with a return

envelope for receipt of the written consent. Both experts were provided with a copy of the kindergarten and first grade case studies. Following obtained written consent a phone conference was scheduled for each expert to conduct the case study review.

*Obtaining consent from focus group-reading coaches.* Each Reading Coach who contacted the researcher was thanked for their interest and was informed of waiting until sufficient numbers were available for conducting the focus group. Upon receiving sufficient numbers, all potential participants were contacted by email to begin identifying a convenient time and date for all. All participants who attended were informed of the purpose of the study and asked to complete a written consent form if they chose to participate in the study.

*Obtaining consent from focus group-specialists.* Each specialist who contacted the researcher either by phone or email was contacted, thanked for their initial interest, and informed of waiting until sufficient numbers were available for conducting the focus group. Upon receiving sufficient numbers, all potential specialists were contacted for the purposes of finding a date and time that was convenient for all who participated. On the day and time selected by the group each member was provided a full disclosure of the purpose and title of the study and asked to provide written consent if they chose to participate in the study.

#### *Researcher Biography*

The researcher's background and biases were identified and documented prior to the initiation of any data collection or analysis procedures to allow for monitoring of potential bias that may influence the data collection and/or analysis phases of the study.

The researcher's biography was also documented to add to credibility to the study (Marshall & Rossman, 1999; Patton, 1990). The research literature used for preparation of the present study provided an initial context for documenting the researcher's biography and potential for bias.

*Background and Experience.* The researcher has worked for several years through graduate training on the use of DIBELS and has been employed within the participating school district as a school psychologist. Both of these experiences have led the researcher to positively value the use of DIBELS and advocate for its use in the schools. The researcher has participated in trainings on the use of DIBELS and has also facilitated training on the use of DIBELS. Some of the earliest training experiences on CBM-Reading, and subsequently DIBELS, occurred at the University of Oregon where the DIBELS was developed. Further background experience relevant to the present context included an advanced practicum project working with representatives from the Florida Center for Reading Research. This experience involved a review of seminal works on reading development and instruction combined with a review of the *Reading First* model in Florida as an exercise in learning how research can inform educational policy. Additional background experiences also involved 4 years of teacher training at a local college on educational psychology and remedial after-school tutoring in reading for students in elementary school grades.

*Researcher Beliefs and Expectations.* The following beliefs and/or expectations concerning the use of DIBELS are provided as the researcher's biases:

1. DIBELS is a reliable and valid assessment tool for screening, progress monitoring, and evaluating student outcomes in early literacy skills.
2. DIBELS, as currently used in the *Reading First* program in the State of Florida uses five subtests which collectively measure three of the five skill domains involved in reading development identified by the National Reading Panel (2000).
3. The PMRN system is useful for tracking student progress and to identify target skill areas for remediation and intervention design for students who are struggling based on the DIBELS results.
4. Interpretation and utilization of DIBELS results requires an understanding of the five skill domains involved in reading development and how they are connected or associated (i.e., phonemic awareness, phonics/decoding, oral reading fluency, vocabulary, and comprehension).
5. There is a difference between screening and progress monitoring.
6. DIBELS is a time efficient, cost-effective, and sensitive assessment tool for monitoring progress towards reading outcomes. “Sensitive,” in this context, refers to the degree to which an assessment tool can reliably identify small gains in performance.
7. It is advisable to collect DIBELS data on a weekly or bi-weekly basis for students who are not meeting benchmark standards defined by FCRR.

Following the identification of the researcher’s beliefs and/or assumptions, a list of hypotheses was developed to provide greater organization and focus to the study for investigation:

1. There exists moderate to high variability among teachers' perceived value of DIBELS in assisting their students' learning needs.
2. There exists substantial variability in the perceptions of non-teacher participants involved in the implementation of procedures for using DIBELS at the school-building level.
3. Given the various other assessment tools used at the classroom level and the overlapping schedules of providing those other assessments (e.g., FCAT, county-wide assessments, Lexile assessments, Running Record assessments, etc.) within a school district, teachers perceptions of using DIBELS are negatively impacted.
4. Teachers understand what DIBELS is and what it measures, but are discouraged or unsure about how to best utilize the data obtained.
5. Finally, given the multiple competing demands and seemingly fast paced nature of school activities, teachers are not accessing their class/student reports on the PMRN, but rather are provided such reports, if any, from the school's Reading Coach or school-based DIBELS team.
6. A low level of direct involvement and use of PMRN reports serves as a barrier for utilizing the data effectively or efficiently.

#### *Data Collection Procedures*

*General procedures.* A variety of procedures were followed regarding the data collection and data management process for the present study. Data analysis procedures are discussed below. However, data analysis procedures were followed concurrently as data were gathered. In general, the following three types of data were collected: Teacher

interview and case study reviews, expert review of kindergarten and first grade case studies, and focus groups with Reading Coaches and specialists. All teacher interviews took place in their respective classrooms after school hours on a date and time chosen by them. Interviews with experts took place via phone calls to each expert, individually, on a date and time chosen by them, respectively. Focus groups took place in a large meeting room located at the participating school district's administration building at a time and date convenient to all participants. All participants were knowledgeable and informed of the researcher's role as a school psychologist and the purposes of conducting the study.

All data were recorded using a digital audio recorder for transcription and later analysis. No identifying information regarding the participant(s) identities were recorded on the digital recorder. All tapes were downloaded onto a personal computer and were coded to ensure that no identifying information was available from their use. Only the researcher had access to the coding system used to identify the tapes. All participants were informed that the tapes and observation notes were coded to ensure confidentiality, and they were prompted when the recording device was turned on and turned off. They were also informed that the purpose of using the audio recording was solely to keep accurate records for later analysis and that only the researcher would have access to the audio tapes. Specific data collection procedures below are provided in order of the presenting research questions.

*Teacher interviews.* All interviews were conducted by the researcher using the Teacher Interview Guide (Appendix C) and the appropriate grade level case study. Field notes were taken at the start of each interview to facilitate later analysis and provide

additional observations not available through the primary interview. Appendix F provides the format used for documenting field notes. Additional comments, observations, and thoughts prompted by each interview were also recorded using the digital audio recorder after each interview for the purpose of adjusting interview questions, for providing context towards later analyses, and for identifying emerging and anticipated topics discussed during the interviews.

The Teacher Interview Guide (Appendix C) was utilized in a semi-structured format (Patton, 1990). The questions used in the guide were derived from a review of the relevant research and information available about DIBELS and the *Reading First* grant in Florida. Any follow-up questions used that were not on the Interview Guide were used as needed to check for clarification and accuracy of understanding of teacher responses. Active search for contrary evidence and alternative perspectives was routinely used during each interview in order to test the limits of participant perspectives and to provide greater reliability. The interview was designed to last 30 minutes.

*Teacher case study review.* Following the use of the Interview Guide, each teacher was asked to comment on their perceptions and thoughts concerning a case study presented to them. Appendix D contains the Case Study Questions and PMRN reports which were used with each teacher for their respective grade level. Teachers were not provided with any background knowledge about the student other than the student's gender, grade level, and difficulty with reading. Questions asked by the teacher concerning additional information about the student's characteristics or history of achievement were captured during the audio recording and written up in the transcripts

for latter analysis. Follow-up questions were provided as needed based on teacher responses to the case study materials. Teachers were generally asked to comment on any aspect they felt was important for discussion and analysis of the case study. The Case Study Review was intended to last only 30-minutes.

Teachers were not initially informed that their responses to the case study would be compared to experts who were provided with the same case study as this would have presented a potential confound to the study. There were no known risks associated with excluding this information from participants. All teacher responses were aggregated for comparison to expert opinion. Individual teacher responses were only used as short direct quotes which highlight a general theme or topic mentioned by many teachers, or which provides a strong example of teacher perception that would otherwise not be conveyed through aggregated or paraphrasing options. In such cases no participant's name, assigned school, or school district was used in any way in the reporting of results in the present study. Teachers were debriefed about the involvement of experts when provided with preliminary results as part of the member check procedure.

After each interview was completed, each teacher was informed of their option to contact the researcher if they had additional questions or comments following the interview. Additionally, each teacher was asked for their verbal permission to contact them via email to check for accuracy of comments and solicit any additional comments based on the overall results. All teachers acknowledge their verbal permission to be contacted again when results were available for comment.

*Expert opinion interviews.* Following consent to participate, each expert was independently interviewed and asked to comment on each of the two case studies through a telephone conference call, one case at a time, using the same Case Study Guide (see Appendix D) that was used for teachers. Follow-up questions listed on the Case Study Guide were used as needed depending on responses made by each expert. Additional questions prompted by expert responses were documented and recorded using the audio recorder for later analysis. The expert interview was intended to last approximately 30-45 minutes. At the end of each expert interview experts were asked for their verbal permission to contact them again if further clarification or checks for accuracy are needed during the analysis phase of the study. Both experts provided verbal approval to be contacted when results were available for comment.

*Focus groups.* The Focus Group Guide (Appendix E) was used during the two focus groups. Questions were developed based upon an understanding of the research concerning DIBELS and knowledge about the *Reading First* grant in Florida. Both focus groups were facilitated by the researcher and assisted by a graduate student with field experience in recording meeting notes identified or addressed during the session. The graduate student also assisted in facilitating a review of comments captured for purposes of validating accuracy of comments and ensuring no additional themes/topics were missed. Consultation between the researcher and the graduate student assistant occurred before both focus groups occurred in order to assist the graduate student assistant in understanding the focus and purpose of the study. Field notes were taken throughout the focus group sessions using a large presentation-style note pad in full view of all

participants. Field notes taken were written as short phrase quotes or paraphrases made by focus group participants.

Each focus group session lasted 90 minutes with the first 60 minutes allocated for focus group discussion using the Focus Group Guide. The final 30 minutes were utilized by reviewing all field notes with the focus group participants for clarification of information recorded and to check for accuracy in the understanding of participant responses. Additional comments or any clarifications made were written down using the large note pad in addition to being collected on the audio recorder. Additional comments provided by the facilitator or in response to facilitator questions not included in the guide were also recorded.

At the conclusion of each focus group, participants were asked for their verbal permission to contact them by email if further clarification or check for accuracy was required as the data was analyzed. They were encouraged to contact this investigator after the focus group was concluded if they had any other comments or ideas they felt would be helpful, or if they had thoughts or feelings that they did not feel comfortable sharing in the group setting.

#### *Data Management and Storage*

The use of a digital recorder allowed all audio recordings to be downloaded onto a personal computer. Each audio recording was coded and saved using a coding system known only by the researcher. All audio recordings of interviews, focus groups, and field notes/observations were transcribed using a computer/word-processing program (i.e., Microsoft Word). All computer audio files and transcript files were secured using a

password only known by the researcher. The elapsed time for each teacher interview was recorded in minutes for each teacher comment recorded to provide a later analysis of duration of time spent discussing a particular topic and/or analysis of percentage of time spent talking compared to the researcher.

Each transcript was saved using the same corresponding identifier as used with saving the audio recordings. Each transcript was then saved as a copy to be used for printing and analysis to ensure the original transcript was not compromised during analysis procedures. All audio and transcript files were backed up on a 4 gigabyte data pin which was secured and kept in the researcher's personal locked file cabinet in the same room as the computer being used for data management and storage. All printed or hand written documents were secured in a file folder and kept in the researcher's personal locked file cabinet. Additional documents produced through data analysis procedures were also saved and stored on either a personal computer and/or in the researcher's personal locked file cabinet.

### *Data Analysis*

*General data analysis procedures.* The procedures for analyzing the data collected in the present study were organized into three different sections based on the three research questions: (1) teacher interviews, (2) expert opinion and teacher case reviews; (3) focus groups. Within each group procedures for organizing the collected data, developing coding schemes, evaluating the reliability of the codes, and applying the coding schemes is discussed. Procedures are then listed next for summarizing and organizing the findings for validation by participants and peer reviewers.

*Teacher interview concurrent data analyses.* The analysis of teacher interviews was conducted in two phases: (1) concurrent initial analysis with data collection and (2) formal descriptive analyses after all data were collected. Concurrent analysis procedures were used with data collection procedures and involved the use of descriptive field notes and personal notes, and a review of completed audio interviews. The focus of this level of analysis provided an ongoing examination of the focus of the study, the appropriateness of questions and instruments being used, and observations of topics or themes which may be important for latter analyses. Adjustments towards each successive interview were conducted based on a review of field notes and personal notes. Descriptive field notes collected were written using a standard form developed by the researcher (see Appendix F). This format was meant to be specific in focus. Personal notes/reflections were recorded using the digital audio recorder and/or written in a journal after each interview was collected.

Field notes and personal notes were summarized using a word-processing computer program and saved in an electronic folder assigned to the participant using a coded label for confidentiality. Hard copy notes taken using the field notes form were secured in the researcher's personal locked file cabinet. The general format of the narrative summary of field notes was in paragraph form to allow for coding at a later time. Each case summary of the field notes taken for a given participant was followed by personal ideas, hunches, concerns, and/or general reflections about the interview/research study. The analysis of personal notes allowed for recording and/or observing of errors and biases.

Review of all field notes and personal notes allowed for the identification of initial topics or themes discussed in order to determine when saturation of important and relevant information was reached. Audio interviews and all associated field notes and personal notes were reviewed in the order they were collected and used to identify initial themes emerging from the interviews. This list of topics was written in the personal notes/journal for later use in the coding process.

*Teacher interview formal data analyses.* After all interviews were transcribed they were separated into one of two groups: kindergarten and first grade teachers and within the order they were collected in order to investigate any salient themes or emphases specific to a particular grade level. All field notes and personal notes were utilized to develop an initial list of topics for each grade level. As each of the teacher interview transcripts was analyzed, additional topics not identified from field notes or journal entries was identified. Both lists of topics from kindergarten and first grade teachers, respectively, were then combined into one comprehensive list of topics (Appendix G1). Numerical codes were developed and ascribed to the initial list of topics. Because the present study involved a relatively tight design – that is, a narrow and specific focus was involved from the start of the study (Miles & Huberman, 1994), and relatively few unanticipated topics emerged from both groups of interviews, the Teacher Interview Guide was used to organize the initial list of codes into different topic clusters (See Appendix G2). Once a comprehensive coding scheme was developed it was applied to each full teacher transcript.

Reliability of the coding process was assessed by using an inter-rater agreement method (Patton, 2002). Inter-rater agreement was assessed by calculating the number of agreements divided by the total number of agreements plus disagreements on 37% of all data (i.e., 6 teacher interviews, 4 teacher case study reviews, 1 expert case study review, and 1 focus group). A minimum of 90% agreement was used as the criterion for determining the reliability of the codes (Miles & Huberman, 1994). Inter-rater accuracy checks that fell below this criterion prompted the research assistant and researcher to review the transcription together and discuss disagreements.

The research assistant who participated in this process held a degree in business administration and had four years experience working as a market researcher. He has experience in conducting field interviews and utilizing both quantitative and qualitative methods for the analysis of consumer perceptions. He did not have any experience working in education. Prior to engaging in analysis procedures, the researcher and research assistant discussed over several sessions the purpose, scope, and nature of the research questions addressed in the study. Hypotheses and researcher biases were shared along with an overview of how educational systems operate (e.g., NCLB, current policies and procedures concerning the assessment and evaluation of student outcomes, *Reading First* model, etc.). All questions asked by the research assistant were answered and discussed until the research assistant indicated sufficient knowledge and understanding of concepts addressed prior to engaging in any analysis procedures.

At the start of the inter-rater agreement process, each code was operationally defined and shared with the research assistant. Two randomly chosen transcripts were

provided to the research assistant (one kindergarten and one first grade teacher interview transcript) along with the list of current codes and definitions. The researcher and research assistant separately coded the same two transcripts and then reviewed each rendition together.

Disagreements found were primarily the result of disagreement over the size of data segments. This led to discussions about decision rules for defining data segments. It was decided that the level of analysis for determining data segments would be at the sentence level. Further, it was decided that each data segment would start with a question posed by the researcher unless the response provided by the participant includes a transition into a different topic. Codes and definitions were reviewed again prior to conducting another round of inter-rater agreement checks using the same (but new/clean) transcripts.

Two additional inter-rater checks were conducted which did not result in reaching the minimum criterion established. The researcher and research assistant discussed at length the disagreements and review of the research questions. Results of these discussions led the researcher and research assistant to conclude three factors were operating which resulted in low agreement percentages: (1) disagreements regarding the identification of prior data segments despite decision rules established, (2) disagreements related to poor differentiation among some of the codes, and (3) disagreements related to an organization of codes a priori using the Teacher Interview Guide.

Following this conclusion it was decided that the researcher should conduct a fresh review of the transcripts and field/personal notes without any a priori search for

topics of interest. The original list of topics developed (See Appendix G1) was condensed into different clusters independent of any a priori organizational structure or topics of interest. This next list was then given a 4-digit code for each topic cluster and applied to all 14 teacher transcripts by writing the code into the margin for each data segment identified.

A manual sorting process (Miles & Huberman, 1994; Patton, 2002) was attempted which involved cutting up and sorting data segments based on those codes. Codes receiving too many data segments were re-read and broken down to provide greater specificity and boundaries in relation to other codes. Codes not receiving any segments were either discarded or clustered with similar codes. This process was conducted several times and involved a high level of resources (i.e., time and materials) and was found to be too ineffective and inefficient for the researcher. At that point, the researcher's doctoral committee was approached to solicit approval for engaging in additional procedures which might prove more effective.

Researchers have recommended the process of coding involve other persons to discuss ideas and define boundaries for topics being identified (Miles & Huberman, 1994; Patton, 2002). The same literature also advocated for the use of computer-based approaches for sorting and coding the data for a more efficient process. Given this review, the research assistant and principal investigator reviewed all 14 teacher transcripts together and identified through discussion of the transcript data where to define the data segments for each transcript. The latest "grocery list" of emerging topics identified without considering a priori topics of interest was then reviewed between the

research assistant and the principal investigator. This process led to the development of a draft of codes. Using this draft of codes, all transcript data for teacher interviews was imported into a Microsoft Excel spreadsheet.

Identified data segments for all 14 teacher transcripts were entered into a Microsoft Excel spreadsheet in order to utilize a computer-based approach to sorting the data. Table 5 in Appendix B contains the format that was used in the entry of the transcript data segments: Topic Code (four digit number), Participant Code (two letter code whereby the first letter indicated the grade level – F or K, and the next letter to identify order of interview – A – N), Data Segment # (i.e., data segments numbered in order and restarted for each individual interview) and finally the transcript content for each data segment with the interviewer listed in italics/underlined for visual ease. After all the data segments for all 14 teachers were entered, the use of Microsoft Excel allowed for sorting the data using the following priorities for sorting: Code, followed by Teacher, followed by Data Entry #. Having sorted the data in this manner, it was found to be more efficient and cost effective for adjusting the codes and developing a third formal set of codes for use (See Appendix G3).

Having defined the data segments together, but prior to engaging in additional inter-rater agreements, a rubric for coding the data segments was developed based upon a priori topics of interest, previous experience with inter-rater checks, and review of transcripts with research assistant. For each data segment, the first step towards coding was to read the researcher's question followed by the participant's response. If the question involved one of several a priori topics used in the study then the data segment

was coded to reflect that topic. The exception to this rule was if the participant's response did not address the researcher's question. For example, if the researcher asked about how often a teacher collected DIBELS data for a student who was struggling in reading, and the teacher either did not understand what the question was asking or described some other aspect of DIBELS data collection that did not address the question asked, then priority was given to the participant's response for coding.

Aside from prioritizing the researcher's question for coding data segments, priorities were also established for which codes to use if there was an option or possibility for multiple codes to be assigned. Of the eight broad topic areas used in the coding scheme (see Appendix G3), priority was given to codes 2000 through 7000 before using codes within the 1000 or 8000 level. For example, if in response to a researcher's question the participant talked about their knowledge of DIBELS measure specifically while also highlighting their perceived value for the DIBELS overall, then priority was given to coding that data segment as "knowledge of DIBELS (2100)" rather than "benefits of DIBELS (1100)." Additionally, if a data segment included multiple topics by a participant and involved a decision between topics listed between the 2000 through 7000 levels, then priority was given to the topic that was closest to the research question asked.

Regarding the use of the "miscellaneous (8000) category, this code was used only for those responses that either involved an exchange of comments between the researcher and participant clarifying the specific question being asked, or reflected responses that were unrelated to DIBELS, reading, or school climate. For example, if in response to a

question about reading assessments a participant discussed concerns with math assessments and student achievement in math, then that data segment was coded as “miscellaneous.”

Though the coding scheme that was developed (see Appendix G3) provided codes for specific examples of responses (e.g., 1310, 1311, 1312), only one of the 28 secondary number levels was used for coding (e.g., 1300). After inter-rater checks were completed utilizing the 3<sup>rd</sup> draft of codes (Appendix G3), inter-rater calculations increased substantially though still just below the required criteria. Analysis of inter-rater disagreements identified discrepancies in coding within the 3000, 4000, and 7000 level topics. Further discussion between the research assistant and the researcher about the nature of the study and its research questions, these three topic levels were each collapsed, respectively, into one level each (3100, 4100, and 7100 – See Appendix G4). It was determined through discussion and reflection that these three topic categories did not require subtle minor topics for coding. For example, the 4000 topic of Progress Monitoring and Reporting Network (PMRN) was a necessary topic because it was an a priori topic of interest in the study. However, too few responses were identified for each of the four sub-categories and the primary focus on this topic was merely on teacher’s general use and perceptions of the PMRN. Therefore, the four sub-categories were collapsed into one overall category.

Observations of using this third draft of codes indicated a higher level of agreement (between 60 and 70%). Discussions of disagreements at this point resulted in some of the higher-order categories being clustered further as described above (see

Appendix G4). This fourth set of codes was then used to transcribe the same two transcripts again which resulted in sufficient agreement (93%). Following this, two more sets of transcripts (four total – two for each grade level) were again used for conducting inter-rater agreement checks. Results of those checks resulted in sufficient agreement percentages ranging between 92% and 97%. Upon reaching sufficient inter-rater agreement on those transcripts sampled for use, the final set of codes was applied to all teacher transcripts using the Microsoft Excel spreadsheet. Again, the data were then sorted and organized by prioritizing Code, Teacher, Data entry #. Following this, attention was given to coding expert/teacher reviews.

*Expert/teacher case reviews.* As with teacher interviews concurrent analysis procedures used during data collection procedures, case reviews conducted by teachers and experts were monitored using personal notes. Reviews of those notes were used for initial formal analysis procedures following the collection of all case review data. First, for each expert, the transcript data were separated into two groups: (1) kindergarten case study and (2) first grade case study. It was necessary to separate the expert transcripts based on the grade level of the case studies for later comparison among teachers for each of the two different grade levels. Each of the two transcripts was read separately for regularities and patterns as well as topics the data cover. Words or phrases used by the experts were used to represent the topics or patterns identified. These words or phrases provided a set of initial coding topics.

Once the list of topics was developed for each expert transcript for each grade level case study, the two sets of topics (kindergarten and first grade case study) for each

expert were analyzed for common and unique ideas in relation to each other for both grade level case studies (e.g., expert one kindergarten case compared to expert two kindergarten case). Common topics or patterns were analyzed first and combined for further coding. Unique topics or patterns were listed together for visual ease and analyzed again for any common ideas or thoughts not originally found in the first steps. Ideas remaining unique between both experts for a specific grade level case study were coded and added to the list of shared common topics for that case study.

The above organizing and coding procedures were also used to develop a set of codes for kindergarten and first grade teacher case reviews. These sets of codes were then used to compare teacher and expert comments in relation to a given case study. However, it was expected that many topics identified by experts and teachers would not be mutually exclusive or unique to either kindergarten or first grade teachers. Therefore, the above procedures provided for an initial comparative analysis between experts and teachers.

A secondary analysis involved the development of a single comprehensive set of codes to be applied to all expert and teacher case review responses. To do this, the topic lists developed for each grade level were compared and analyzed for common and unique topics. Common topics between the teacher cases were combined and reorganized. This list was then compared with the expert list of topics and a constant comparison procedure was used to develop a final comprehensive set of topics to be applied to both expert transcripts and teacher transcripts, regardless of case study discussed. This final list of codes involved three distinct topic areas for coding: (1) Using PMRN reports; (2)

Additional information needed or wanted; and (3) General comments about using DIBELS data. A computerized method for organizing and coding the teacher/expert case reviews was not needed for the analysis of case review data.

Inter-rater agreement checks were applied to two sets of Teacher Case Study Reviews (Total of four – two for each grade level), and one of the Expert Case Study Reviews (for both kindergarten and first grade cases). As with teacher interviews, a 90% agreement criterion was used to check inter-rater reliability. No additional procedures were needed to secure sufficient agreement between the research assistant and researcher. Inter-rater agreement percentages for all case reviews (both teacher and expert) resulted in 100% agreement.

*Focus group analyses.* To begin, the Reading Coach focus group and the specialist focus group were initially separated while the following procedures were applied to both groups. Prior to reviewing each focus group transcript, notes taken and validated by focus group participants was reviewed and an initial list of topics was identified. The first list of topics was then kept aside while the focus group transcript was read for topics and themes not captured in the notes taken. A review of both the focus group transcripts did not yield any new topics beyond the notes taken during each of the focus group sessions.

The first formal draft of topics for focus group data was developed by engaging in a constant comparison method (Patton, 2002) in which similar topics were combined and unique topics were listed without exclusion. From this process a comprehensive list of topics was identified for the focus group data. Each topic was assigned a number and

applied to both focus group transcripts. No additional procedures (i.e., computerized approach to organizing and sorting the data) were needed for coding the focus group data. Inter-rater agreement procedures were applied to the Reading Coach focus group transcript. Results of the inter-rater agreement yielded a satisfactory agreement percentage of 96%.

*Member checks and peer reviews.* In preparation for evaluating the validity of findings generated to this point, the coding systems developed for teacher interviews and both focus groups were compared using a constant comparison approach (Patton, 2002). A comparative review of these sources of codes revealed a pattern that prompted a further organization of topics using the following four categories: Climate/Culture of School, Supports/Resources Available, Knowledge of DIBELS, and Collecting and Using Data. Table 6 in Appendix B depicts the consolidation of topics into these emerging final four themes. Preliminary results were written using this organizational format using an outline version for ease of reporting and reading (Hodges, personal communication 2008). A copy of the preliminary results (See Appendix H) was developed for each group of participants whereby the section reporting their group's data was highlighted in yellow for ease of review. Each participant in the study was provided via email a copy of the preliminary results and asked to provide feedback on the accuracy and completeness of the information.

Peer reviewers were given the same report without highlighted sections. All versions included a table of contents to assist participants with an efficient means for finding areas of interest. The preliminary results were emailed to all participants and

asked to respond by emailing the researcher within two weeks from the date the report was sent. Participants were asked to comment on the findings and were encouraged to offer any additional thoughts prompted by the results. Peer reviewers were asked to review the results and offer any critical feedback or questions concerning the findings, the organization of reporting, and any other methodological aspects. Responses received were used to further analyze the validity and reliability of the study and are reported in the results section.

## CHAPTER FOUR - RESULTS

### *General Introduction*

Given the methodological design used in the present study, the results were written in two parts. First, in accordance with the research questions posed at the start of the study, the findings obtained in relation to those questions are presented. Next, an analysis of the research process and the researcher's perspectives throughout the course of the study are presented. The primary findings as they relate to the research questions were written in a traditional third person voice. The researcher's analysis of the research process and personal accounts of the evolution of the study were written in first person voice.

### *Description of Participants and Data Obtained*

Procedures for selection of participants and access to sites were provided in the previous methods section and therefore are not restated in full here. However, it is important to highlight that all participants self-selected themselves for participation in the study. All participants were asked to take part in the study through a one-time interview or focus group session. A total of 831 minutes, or nearly 14 hours of audio-recorded information was obtained directly from participants as a whole. When transcribed, these data amount to 229 pages of text.

Aggregated teacher interview data may be described as 381 minutes of one-on-one interview discussions or 135 pages of transcripts. The average participation time in

the teacher interview was approximately 27 minutes with a range of 13 to 43 minutes. A total of 617 separate data segments were identified for coding/analysis purposes. Teacher case study review data may be described as 141 minutes of audio recording or 44 pages of transcribed text. The average amount of time spent discussing the case study review was approximately 10 minutes. Expert case study review may be described as a total of 47 minutes of audio recording or 10 pages of transcribed text. Focus group data may be described as 162 minutes of audio recording or 47 pages of transcribed text.

*Teachers' Perceptions and Understandings about DIBELS and the PMRN.*

*Climate and culture of schools.* Several a priori and emerging topics were analyzed in order to understand teacher's perceptions and use of DIBELS and the PMRN. To begin, the culture and climate of education within which teachers had been asked to work emerged as an important factor to understand. Overall, teachers reported experiencing various levels of pressures and stress related to their profession. Most of these comments seemed tied to a general climate of accountability to increase student performance in education. Many sources of pressures were identified through teacher interviews. Some of these pressures were reported to have evolved in relation to the implementation of *Reading First* and the use of DIBELS. Some of the variation in reaction to accountability pressures was found to be related to the school's status or grade as evaluated by the State of Florida's Department of Education using the No Child Left Behind criteria as one teacher indicated.

*So, we've been under a lot of pressure because we are not making A.Y.P.; Annual Yearly Progress. We are one year away from restructuring. And, I work*

*extremely hard, and every person on this campus works extremely hard. So the feeling is there across the board. It comes from No Child Left Behind. To expect every child to be on or above grade level like 2012 is the most ridiculous thing I've heard in my life. Unless you are cloning the children, it's not going to work because children learn in different stages.*

A great amount of pressure was expressed by some teachers as a function of the DIBELS test being implemented into the schools. Most of the teachers reported an initial impact described as increased competition among teachers. However, this initial impact of DIBELS had evolved in most of the schools sampled as teachers reported they no longer perceived anyone would lose their jobs due to the outcome data on the DIBELS. They reported that collaboration as a grade level team was functional for determining what was working or not working for their students. When asked to generally share her views about the DIBELS at a *Reading First* school, one teacher stated the following:

*We've been doing it for a few years now. When we first started none of us knew what it was about or how it would impact us. And then it was kind of a competition between teachers; you know whether your scores went up or your scores were down. And then we decided it was more of a team thing because it depended on more than just your class. Sometimes you might have a really young class and so then you just kind of looked at your scores and how they were improving and not just where they're at but where were they and where are they at now?*

In general, all the teachers interviewed expressed believe in the importance of the role of leadership at their school as it related to pressures and use of DIBELS. The role of leadership was reported as important in terms of setting a climate of cooperation and a culture of using data to make decisions. It was also described by teachers as important for ensuring that teachers use the DIBELS as the *Reading First* grant intended. Most reported that DIBELS had been given a greater value by the school leadership or school district than other assessments. And in some cases, uncertainty was reported about why; as if the DIBELS was the main indicator for student progress at the exclusion of using multiple sources of data. Or, as one teacher commented, “There’s a lot of expectations and I really think that sometimes it’s more important than the classroom assessments. Because I do think they feel that this is the way to see where these kids are coming from and where they need to go, and how you’re gonna get ‘em there.”

Another teacher described the importance of school leadership in helping to deflect outside pressures of accountability while another teacher described the isolated feeling of working in a school that is very “political.” Several of the teachers commented on the concern and simultaneous pressure to “teach to the test.” But over time, influenced in part by school leaders (including the Reading Coach-see below) and increasing student achievement scores, some of these teachers reported less pressures about how their student’s were performing on the DIBELS and instead were more encouraged to use the data to help support student learning.

Aside from perceived pressures brought about by supervisors, some teachers described their concerns about the general climate of their educational field as “removing

the childhood from the child.” It was reported by some of the kindergarten teachers that there was too much intensity in kindergarten compared to the past. This view was further characterized by teachers as if what kindergarten students are being asked to do is developmentally inappropriate. When these teachers were asked if that was indeed their view, they all indicated an affirmative response. One teacher also characterized her view as too much emphasis in reading at the expense of other kindergarten goals such as social-emotional development. As she stated, “...yes, they get real strong on the DIBELS but their behavior hasn’t gone to a place that is appropriate for society.” When asked to further explain her point, this teacher modified her response to reflect a more general feeling that an intense focus on reading was making some teachers feel less and less time was available for helping children learn to develop positive relationships and social skills. A different teacher commented on this similarly when she spoke of her fears regarding a rumor that the district would be considering the removal of rest time for kindergarten – a time she valued highly both for her students and for her to catch up on classroom assessments and paperwork.

Perceived pressures were also reported to be occurring from other grade levels. One teacher described the pressures felt by third grade teachers in relation to the FCAT and No Child Left Behind as being passed down to the second grade teacher to make sure that exiting second grade students were prepared to move into third grade. In turn, the second grade teachers expected the same from the first grade teachers and so on. The teacher who presented this view also indicated that this was often a contributing source of why retention was considered at the grade levels below third grade, especially first grade.

A final source of perceived pressure identified among teacher responses was the concern over the amount of testing that was occurring in the schools. In identifying the various reading assessments that can occur at a *Reading First* school in the school district that participated in the study, Table 7 depicts the assessments and frequencies of administration were reported by teachers – in some cases the specific name of the test was withheld in order to maintain confidentiality of the school or district. The frequencies of administration alone throughout the year were deceiving as, for example, the kindergarten district assessments were conducted in a one-on-one format between teacher and student. Kindergarten teachers indicated, on average, that it took about 3 weeks to complete a whole class using that assessment because of its required one-on-one assessment format.

Some teachers shared their concerns over the loss of instructional time due to the amount of testing. One teacher stated it in the following way which was very representative of kindergarten teachers in this study: “That’s all we’re doing is testing. They’re using it for that reason, but I want to teach the children. There’s just no time to do that. And, um, so that’s what drives a lot of teachers crazy.” And another teacher took a more humorous approach by saying, “Sometimes I look at them and think, how did I expect you to get...to go higher when I haven’t had any instruction time in between. It’s a miracle!” Compounding this stated teacher frustration, some reported frustration with having to conduct assessments that were not useful for driving instruction such as the district’s school readiness test; a test that one teacher wished the district would

arrange for students to take before the start of the year because of how little value it provided her and the amount of time it took to complete a whole class.

It is relevant to note in a context of pressure and accountability that all teachers reported concerns about using the DIBELS data more effectively and efficiently as a result of not being the ones trained to administer the test. Only two teachers indicated an interest in further training to learn how to give the DIBELS themselves; one other teacher had already been trained to give the test. All the others cited the current assessment responsibilities they have as a barrier to taking on the administration of the DIBELS. And yet, these same teachers indicated a preference for using their district assessment for monitoring student progress between assessment cycles rather than using DIBELS because (1) they had access to the materials they needed and know how to administer the test, and (2) they valued being able to directly hear the student read in order to grasp a qualitative understanding of what the student can do. One teacher stated it in the following manner when asked if it would be beneficial to learn how to administer the DIBELS herself in order to have more direct observation of student performance:

*It might make a difference in some cases. But, we're talking about timelines here and individual one-on-one testing. So how they would ever get it so your class is covered and the teacher does all this testing? Because if one person had to test 20 kids, that would be a lengthy process.*

In response to a question about what would be needed to sustain the use of the DIBELS assessment when the *Reading First* grant is terminated, almost all teachers indicated that the school would need additional personnel to help with collecting

assessments – not just DIBELS. In fact, as it stood when these teachers were interviewed, many of them had concerns about the personnel that were then involved with helping to collect the data (i.e., Title 1 staff), as doing so meant missed instructional time for students with those staff members.

*Support/Resources for teachers use of DIBELS.* Having provided a description of the school climate/culture and expectations perceived by teachers, analysis of the first research question then focused on the topic of supports and resources that existed in schools to help teachers with their understanding and use of DIBELS. Consistently, across all teacher participants, the *Reading First* grant was praised for the various resources it provided to schools. Teachers reported high value for the grant and its provision of books, curricular and instructional materials, and additional media materials. But the most valued of all the resources provided to schools, as reported by teachers, was the assignment of a full-time school-based Reading Coach.

Teachers expressed numerous reasons for their high value for the Reading Coach. According to the teachers interviewed, the Reading Coach provided coordination of all data collection and data input efforts. The Reading Coach was reported to provide technical assistance to teachers on the use of DIBELS and other assessment data for making instructional decisions. The Reading Coach was reported to help some teachers create and /or organize various activities for their classrooms as well as help teachers organize their classroom libraries. Some teachers reported receiving frequent support from the Reading Coach on the use of the Progress Monitoring and Reporting Network (PMRN) reports– generated from a web-based program that provides educators with

graphs of their student's DIBELS data. More specifically, Reading Coaches were reported to meet with grade level teams at least after each DIBELS benchmark assessment to assist those teams in using the data to develop action plans for increasing student reading performance.

Specifically, with regards to support given by the Reading Coach for using the PMRN, teachers reported they highly valued the accessibility of color graphs which were provided to them by the Reading Coach. The PMRN reports utilized a color coding system to assist teachers in helping to scan their classroom data and identify areas of instructional need. All teachers indicated that they did not have access to color reports and were dependent on the Reading Coach to provide them because of an insufficient supply of color printers available on school grounds. Some teachers, who valued the use of the PMRN reports, indicated that when they did download and print their own reports in black and white, they were left with having to color them in by hand for visual analyses.

The lack of access to color printers may have been a potential barrier to teacher's use of DIBELS data, let alone a barrier to teacher's use of the PMRN (discussed in more detail later). Nonetheless, teachers voiced much appreciation for their Reading Coaches who at times also provided modeling of instructional programs or strategies. And in some cases, teachers reported their Reading Coach had set up site visits to other schools for them to observe expert teachers who utilized a given reading program or instructional approach that the Reading Coach was interested in helping her team learn. One teacher's

view in particular may best convey the general sense of teachers' valuing the Reading Coach:

*Um, she came in...for a six week stretch and modeled lessons for me. We co-taught some lessons and I did some on my own and she observed and gave me feedback, that kind of thing. Looking at the differentiated instruction and making games up for rotations, and things along those lines, and word work. And um, she's always been there as a resource. Questions me...you know, helps me think things through cause I don't see it somehow, she's been there in that regard as well. Um, working with getting the classroom library up and going. The whole program started with the grant and they bought all those extra books and we've been doing that. Wonderful person. I mean, without her I don't think it would have went as smoothly? Yes! Definitely not.*

The topic of classroom libraries and books was another major contributing factor influencing teacher's value for the *Reading First* grant. Most of them indicated having access to "leveled books" as part of their personal classroom library for use in their reading center activities. The classroom library was often mentioned in the same context as providing students with a differentiated instructional approach that allowed all students to work at their instructional level – an instructional design being promoted through the *Reading First* model. Prior to having the grant, teachers reported they had few other options for supporting their classroom libraries and similar curricular or instructional needs than to use their own money for purchasing such materials. In this regard, the grant was viewed by the teachers in this study as highly beneficial.

A less common positive impact of the grant reported by teachers was the introduction of DIBELS. Some of the teachers (4 of the 14) indicated a direct relation between use of the DIBELS and the increases in their overall school performance. Most of the teachers, when discussing the impact of the grant, merely described DIBELS as a requirement for which they had mixed views. Or, as one teacher commented, “It’s just another test.” But a few teachers commented on how DIBELS had been a very valuable resource provided through the *Reading First* grant because it helped them to focus their instructional efforts and guide their professional development in reading. More specifically one of these few teachers stated, “Um, I was resistant in the beginning because of the amount of time. I see some good results with DIBELS and *Reading First*...What has happened is that it has created an awareness of what’s important. I think that’s the big thing of value.”

A final area of support identified through teacher interviews was the availability and focus of trainings provided to teachers on the use of DIBELS data and the PMRN data reports. None of the teachers reported attending any district level trainings on the use of DIBELS data. However, some of the teachers reported attending a mandatory training provided to the whole staff at the start of the first year of implementation. This particular training was described by these few teachers as having low value as it was reported the training consisted of a facilitator reading from a scripted manual over several hours of training across four days.

One teacher indicated she had attended a district training to learn how to administer the DIBELS. This teacher indicated the following in response to a question about the general atmosphere among school staff on the use of DIBELS:

*“I think people don’t appreciate the DIBELS that we have. I don’t think they know enough about it to appreciate the information that they’re truly getting. For a lot of people, it’s just another test. Um, for the new teachers that have come in, as team leader, I’ve really tried to impress and help them to understand the information that we are getting. Um, as a first grade team I think we understand it, but some of us still don’t appreciate it; at least some of the older generations.*

Although all of the teachers indicated receiving training either as a grade level team or as an individual from their Reading Coach, most reported those trainings consisted of how to use or implement various intervention/instructional ideas in relation to the DIBELS data. Few teachers indicated any participation or formal training in how to use DIBELS data to inform instruction – at least to a level of independence without assistance from the Reading Coach. Most of the teachers reported receiving some amount of training on how to interpret the colors of the PMRN graphs to identify students at-risk for reading difficulties and for determining what students should receive additional instructional services (e.g., use of specific curricula or participation in pull-out programs). Regardless, it was difficult to determine whether the continued dependence on Reading Coaches to use DIBELS for instructional decision-making was a function of a lack of sufficient training, or a direct result of teacher’s reluctance to embrace such activities in light of the

competing responsibilities described above with assessments and the pressures over student performance.

Some of the teachers, however, reported an interest in further training. These same teachers had also voiced positive uses of the DIBELS assessments. Teachers who perceived little value in the use of DIBELS declined to participate in any further training. Of the teachers who indicated an interest in more training on the use of DIBELS data, two of them indicated a specific interest in using DIBELS for more frequent progress monitoring among their students. Those teachers who indicated an interest in more training stated they were interested in not only how to administer the DIBELS, but to learn more about its development as a test and how to use it more effectively to guide instruction. An example of the kind of information that teachers find valuable was provided by one teacher who indicated learning about the high correlation between the DIBELS Oral Reading Fluency subtest and the FCAT.

*Knowledge of DIBELS.* Teacher's knowledge of the DIBELS was a primary focus of the teacher interview. Identifying what teachers know about the DIBELS was expected to provide information to help understand how teachers perceive and utilize the DIBELS. All teachers indicated knowledge of the frequency of DIBELS testing as a benchmark assessment – three times a year. This frequency just recently changed at the time of the interviews from four times a year during the prior school year to three times a year. All teachers reported knowledge of the use of the DIBELS as a required assessment through the *Reading First* grant.

As interview questions explored more specific knowledge about DIBELS, substantial variation of responses was found. Such topics explored included the timing of students, the perceived correspondence with other assessment in use, knowledge of specific subtests and what they measure, and knowledge of DIBELS as a progress monitoring tool.

*Timing students.* Half of the teachers who were interviewed voiced concerns over the timed aspect of DIBELS. Of these seven teachers, five were kindergarten teachers. Kindergarten teachers in particular were most vocal about this concern. Some of these teachers felt the timing of students was unnecessary and/or inappropriate. A representative view of timing students in kindergarten was stated in the following way:

*Well, one of the things I find, um, difficult from the children's standpoint is that they're timed at this early age. Especially at the beginning when they don't even know what letters are; I mean for the most part. And uh, you know, to put a test like that in front of a child seems intimidating to me. And so I don't like that very much; I don't like the timed aspect. I mean, and that's what it is, it's a timed thing. I feel like it's a little bit pushy.*

These seven teachers who voiced concerns about timing students perceived the DIBELS test as a kind of “speed test.” Further, some of these seven teachers perceived less reliability or validity of the DIBELS because it involved a timing aspect. And yet, the other seven teachers indicated a perceived higher reliability in DIBELS because it was a measure of student’s fluency of a particular reading skill – often discussed in the context

of comparing it with Running Record assessments (for first grade teachers) which were described by these other teachers as a measure of reading accuracy only.

The perceived value of DIBELS as a “speed test” among the seven teachers who voiced concerns about timing students was also seemingly impacted by these teachers’ concerns that reading fast was not the goal of reading; rather it was to understand what they are reading. A first grade teacher stated it in this way which was characteristic of the other teachers sharing this view:

*Running Records, um...I don't time them, but if they are reading at a nice talkable speech, then I think that's fluency. And I encourage them when they stop and have to sound out, that's exactly what I want them to. They don't have to read like an eighth grader or a tenth grader. We talk about how we read as we speak and I talk about the wind and the waves going back and forth, and that's when not-to-talk-like-this (imitating a robot voice).*

The teachers who identified concerns about timing students highlighted their concern about perceived importance being placed on how fast kids read rather than their level of comprehension. These teachers stated a preference for students reading slow but understanding what they read over reading fast and not understanding what they read.

Or, as one teacher said, “I’d rather have a slow reader comprehend like that (snaps finger) after talking to them which is what I as his teacher could do, versus DIBELS which is...so you read 23 words and the criteria is 40 words; uh, you have no oral reading fluency. He does have oral reading fluency; he’s just a slower systematic reader.”

Some teachers who voiced concern over timing students perceived students feeling anxious and nervous about being timed. And yet, when asked if any students have ever voiced such concerns directly, none of the teachers could recall a specific instance of students voicing concern about being timed. But teachers themselves perceived pressure and stress over the timed aspects of DIBELS to warrant a view that students must themselves have similar feelings. For example:

*(Researcher): Have any of your students ever brought up any concerns about not knowing someone or feeling inhibited to perform or talk?*

*(TCH:) They.... No, they haven't brought any concerns up, but I know...just as a teacher, when they go to...I know they're...even just bringing them into another classroom they're like (models child behavior by making stiff posture with body and avoids eye contact). They just change... They just change; they're quiet, they're...you know...I mean I could have a behavior problem over there and bring them over here and their behaviors change just 'cause he really don't know [the other teacher], or who ever else...you know, they don't know so they're a little bit more withdrawn than when they are with their own teacher.*

Additionally, concerning student's reactions to being tested with the DIBELS, all teachers perceived the potential for unreliable data if a student is unfamiliar with the person who is conducting the assessment. Though this concern was evident in all teacher responses, some indicated a greater concern than others. Some teachers reported this as a concern primarily in the first cycle of assessment; after which students were reported to adjust and there was then less concern about this topic. But other teachers maintained

that even with knowing or being familiar with the person conducting the assessment, they perceived students were likely to not respond as well as if they were performing for their classroom teacher. It was noteworthy that the teachers most concerned about students being tested by persons other than the teacher were kindergarten teachers who reported low value for the DIBELS. It is difficult to determine if teachers' perceived low value for the DIBELS led them to perceive anxiousness in the students or if their perception of student anxiety contributes to their low value for DIBELS.

Some teachers reported procedures used at their school to help students understand the expectations for participating in DIBELS assessments. Generally, a person, such as a Reading Coach, would visit the classroom, either a day or two in advance, or on the same day of testing to talk to the students about the process and to describe what they would be asked to do. Some teachers reported the use of stickers and similar rewards given to students following their testing experience. Only one teacher actually reported positive value with the DIBELS being tested by someone other than herself as she found it reassuring to know that someone else was observing students' successes or difficulties; thus potentially validating her own observations in the classroom.

Teachers who did not voice any concern about timing students generally reported DIBELS as a valuable tool for identifying students who need help in specific areas of reading. Though these seven teachers generally understood the DIBELS as a measure of fluency, some of these teachers also perceived less value in the DIBELS as a test more focused on how fast students read rather than understanding what they read (i.e.,

comprehension). This perception is discussed later in the context of understanding what DIBELS measures. When asked about students' reactions to being timed, some of these seven teachers reported it was the teacher's responsibility to set the expectations for students by preparing them for assessment as well as, "making a game of it." For example, one teacher indicated she encouraged students to do their best by competing against their own scores. This same teacher also had incorporated her own version of progress monitoring of all students in the classroom in addition to teaching her students to graph their scores (discussed further later).

*Correspondence with other assessments.* Aside from timing of students, teachers were asked to comment on their perceptions of the DIBELS as compared to the other assessments they were using in the classroom. The intent of exploring this topic was to better understand teachers' perceptions of how to utilize the DIBELS assessment in relation to other assessments to determine student needs. Among the 14 teachers interviewed five of the teachers voiced preference for classroom assessments over the DIBELS or, in all five of these cases, perceived little value for the DIBELS because it did not correspond with other assessments (four of these teachers were kindergarten teachers). Two other teachers reported mixed views on this topic; indicating they generally perceived a correspondence between DIBELS and their classroom assessments, but they still preferred to use classroom assessments to guide instruction. These teachers reported the data from these classroom assessments were more immediately available for use since the teacher was the one giving the assessment. The remaining seven teachers all reported value in using DIBELS in conjunction with their classroom assessments. For

example, one teacher had the following to say when asked about any concerns using the DIBELS with other assessments:

*(Int:) Ok, and do you find any difficulties or concerns with making use of the DIBELS information along with these other assessment options?*

*(TCH): No, um, I think it all pretty much coincides because if you know that their reading fluency is low then their Running Record is probably not going to be anywhere near a 17 which is what it needs to be at the end of the school year.*

*Um, if it shows that they're, you know, in that yellow area they're probably in the process of learning. I think what you do in the ORF (oral reading fluency) directly correlates with what you do on the Running Record.*

One of the 4 kindergarten teachers who perceived a lack of correspondence between the DIBELS and her classroom assessments cited concerns over how the student's demonstrate knowledge of letter names through their classroom assessments, but then test poorly on the Letter Naming Fluency subtest on the DIBELS. Though specific to a particular DIBELS subtest, the general theme was consistent between these 4 teachers. It is notable that these same teachers also voiced concerns about timing students on the DIBELS and did not perceive DIBELS as a valuable measure of reading ability. Thus, it appeared that the perception of DIBELS as a "speed test" was associated these teachers' perceptions about the utility of DIBELS in relation to their classroom assessments.

When teachers were asked if the DIBELS provided any information that their current classroom assessments did not, some teachers – even some who found

correspondence between DIBELS and the district assessment – reported a low value for DIBELS citing no new information available through the use of DIBELS aside from what their classroom assessment provided. However, some teachers who valued using DIBELS with their classroom assessments perceived relative benefits in using DIBELS over their classroom assessments. These teachers were first grade teachers and commented on DIBELS in relation to the use of Running Record assessments; they generally reported the DIBELS to be more useful for analyzing a student’s errors in order to determine instructional ideas. These teachers also cited how the use of Running Records “just let’s you know if the student can read.”

All teachers reported that they preferred the DIBELS over other assessments in being more time efficient to complete a class of students. Different schools indicated different procedures for collecting the DIBELS, but two main approaches were identified. One was to schedule classes to visit the media center where the students would be occupied with an activity while students were assessed by one of several testers. The other procedure was having each of the DIBELS testers select classrooms/grades to pull students from one-by-one to a nearby area to complete the assessment. Regardless of the approach used at a particular school, all teachers indicated the amount of time needed to complete a class was around 20-40 minutes. The DIBELS was further valued over classroom assessments in this context because teachers were not the ones testing the students and could instead focus their time on instructing their class. Kindergarten teachers reported that it could take as much as three weeks to complete their whole class using the district assessment – during which time they found difficulty in trying to

provide instruction and collect assessment data simultaneously. First grade teachers reported similar views in the use of Running Records.

On the topic of DIBELS knowledge, five of the teachers interviewed expressed concerns about the use of the Nonsense Word Fluency (NWF) subtest which was an unanticipated topic by the researcher. Two of these teachers had perceived positive impacts on the use of DIBELS at their school, and yet all 5 of these teachers viewed the NWF as highly confusing for students, incompatible with the strategies they are teaching in the classroom, and less valued because it asks students to read “silly” words, as one teacher called it.

Among these five teachers, the NWF subtest was seen as very confusing for students because they were taught specific strategies in the classroom about how to identify a word if they don’t already identify it by sight memory. The general strategy reported among these teachers was to encourage the student to look at the word, sound out the first part or all parts known, and if the word doesn’t sound right, then think of words they know already that would fit or make sense. Additionally, some teachers reported they encouraged students to use picture clues and context to identify words they do not know or cannot read. Given this, these teachers believed it confuses a student to learn a particular set of strategies for deciphering an unknown word and then have to “throw out what they have learned” in order to perform on the NWF subtest. A couple of teachers were so passionate about their concern over the use of the NWF subtest that they wanted district leaders to throw the whole NWF measure out because, “We don’t teach nonsense words in kindergarten!”

Two of the remaining nine teachers had no comment either way about the use of the NWF subtest. However, the remaining seven teachers identified the NWF subtest as a measure of student's decoding ability – not their sight word reading ability. Regarding the teachers who perceived no value in using the NWF measure, when further interviewed about their knowledge of the NWF measure, some misunderstandings were identified about its use as a measure of decoding skills. When asked this question directly and why some teachers do not like this subtest, one first grade teacher who valued the NWF subtest stated, "I've heard that from the DIBELS assessors. And yet on the PMRN parent letter it says that the nonsense word piece is in there so that we're not assessing the reading ability but the decoding ability." However, one teacher who did have an understanding of the NWF subtests still voiced concerns regarding its use.

*Um, I'm sad for my kids when they become such a great reader and they forget how to read those words or it takes them so long because they want to try to sound it out and make it into a real word. I think it's a very useful tool because you know what these kids know. But, um, we work a lot on nonsense words in here as far as our word work is concerned. Um, we use dice and we roll them and talk about what a nonsense word is. And I want them to be familiar with that concept which is why I do that work with them. But as they get into higher Running Record, they almost try to put too much meaning into it. And I think in that aspect it makes it harder for these children.*

This teacher had been incorporating direct teaching of decoding skills with the use of nonsense words in the classroom while also generally advocating for its use. And yet,

she indicated concern over how students will sometimes struggle on that particular subtest. From her perspective, more specifically, she did indicate that this observation mostly occurred with students who were already reading above grade level standards on the Running Record assessments in her class.

The teachers who valued the use of the NWF subtest, however, reported that the score alone was insufficient for them to assist students in their decoding skills. These teachers all generally highlighted the use of the actual protocols used to record the students' performances during the assessment. They indicated a value in being able to identify the patterns of errors using the student protocol to determine the appropriate focus of instruction. As one teacher put it, "...it's the only way to really see what you're actual scores show you in the end. And so, just because they tell you they're red, green, or yellow, or blue, doesn't mean anything unless you get to see those tests along with it."

Additionally, some teachers indicated high value in the use of qualitative notes written on the test protocols by assessors to further aid teachers in the interpretation of scores (e.g., notes that the student wasn't feeling well.). However, teachers reported high variability in the quality of the notes shared by DIBELS assessors on the protocols. Thus, at times the notes were reported to be very helpful, and at other times, not so helpful or simply not provided. In this context, a couple of teachers did indicate some interest in receiving training on how to administer DIBELS in order to see for themselves directly how students respond on the DIBELS.

In fact, many teachers, when talking about their own district or classroom assessments valued giving a test themselves because they valued the qualitative aspects

of hearing the students read for them. And yet, when asked if they'd like to learn to give the DIBELS in order to maintain the same opportunity to directly view student responding, most declined by citing their current assessment responsibilities. At the time of the present study, no information was available regarding the reliability and validity of the district assessments in use for kindergarten and first grade. One must then consider, would it be appropriate to trade district assessments in for an assessment or set of assessments that are less distracting from instructional time and yet offer greater reliability and validity for classroom decision-making?

On the topic of specific subtest perceptions two teachers did voice concern over the use of the Phoneme Segmentation Fluency subtest. These teachers believed it confused students when they are being taught to blend sounds together to make words and yet on the DIBELS PSF subtest, they were being asked to separate sounds from words. It seemed like these teachers may not have understood the purpose of the PSF task – to assess an auditory skill. Segmenting and blending skills are both important auditory early literacy skills for reading (National Reading Panel, 2000). It should be noted that these teachers had also expressed concern over students being asked to read nonsense words. Other teachers, however, voiced positive value in the use of the PSF subtest because it was measuring similar skills to either what was measured in the district assessment for kindergarten or because it helped the teacher to determine if the student has mastered their phonemic awareness skills.

One teacher in kindergarten voiced a concern over the use of the Initial Sounds Fluency subtest. This subtest presents a student with four pictures and the pictures are

labeled for the student by the assessor. The student is then asked to identify a picture that begins with a specific initial sound and the latency in their response is measured.

Specifically, this teacher cited concerns over the use of the labels given to a particular picture. She indicated that many students who speak English as a second language may use a different word for a given picture. It is important to mention that this teacher does value a measure of students' abilities to segment the initial sounds of words, but specifically questions the labels used in the assessment for particular pictures.

*And it's not that the actual procedure – but that the pictures are bad. For instance there's one, and I don't exactly remember what it is, there's a house and a road, and there's some grass, and there's supposed to be this thing, a yard. Well, you know, they might not even use yard or they might use yard. Another word is luggage. Um, one of them either city or buildings and it's like...it's just a language thing. So the child has to go what did they call [it] this time? And I practice them with my children. This time they might call this a puppy. This time they might call this picture a dog. So you're going to have to listen to what they call it so that you will know how to answer. It's very hard to remember. They're just learning the language.*

Regarding the Oral Reading Fluency subtest, first grade teachers offered mixed views on this subtest. Again, one of the variables leading to the mixed findings seemed related to the issue of students being timed. A first grade teacher who held very strong concerns about timing students, and whose perceptions were unique, indicated concerns about the criteria for first grade students at the end of the year. She felt that it was too

high for them and questioned the validity of the subtest as a “speed test.” Yet, other first grade teachers perceived it as a risk-indicator of being able to comprehend what they are reading. One first grade teacher, as mentioned previously, had learned that a correlation had been found in the research between the ORF score and the FCAT. Some teachers reported some concern about how to interpret a student’s high performance on the Running Record assessment and a low performance on the ORF subtest, while other teachers did not have this problem as they perceived the Running Record assessment as a test of accuracy and the DIBELS ORF as a test of fluency.

*DIBELS and progress monitoring.* On the general theme of understanding teachers’ knowledge of the DIBELS, the issue of using DIBELS as a progress monitoring tool was investigated. All teachers were asked if they used DIBELS for progress monitoring more frequently than the three times a year for benchmark purposes on students who are struggling in reading. All teachers indicated they did not, but the reasons for such were varied. For some, it was a matter of having access to materials for use – citing knowledge only of the materials that were available for benchmark assessments. For others, it was a matter of who would collect such information as teachers were not trained to give the DIBELS. For others, it was not valued to use DIBELS for progress monitoring as they voiced preference for using their district or classroom assessments/observations to track student progress. The reasons cited by teachers for this particular preference seemed to converge on teachers’ comfort in observing their students reading directly in order to observe what they can or cannot do.

However, one unique teacher who was interviewed did indicate developing her own version of progress monitoring materials, using a one-minute time limit, in order to track what students were learning between benchmark cycles. This teacher also indicated teaching her first grade students how to graph their performance on small index cards while encouraging them to often practice and “try to beat their own score.” Two other teachers reported an interest in learning how to use DIBELS for progress monitoring citing barriers to do so because of a lack of access to necessary materials. When learning about the option to receive such materials on-line through a companion website for DIBELS, these teachers reported interest in learning how to administer the DIBELS for such use. None of the 14 teachers indicated any knowledge of the option to download and print out DIBELS progress monitoring materials from the official DIBELS website developed through the University of Oregon. The remaining 11 of 14 teachers, upon being asked if they’d be interested in learning how to administer the DIBELS for progress monitoring purposes, reported a feeling of being overwhelmed with the current level of assessment responsibilities charged under their care.

*Using assessment data.* To this point, an understanding of the climate and culture of education in which teachers are working, identification of what resources exists to support their use and understanding of DIBELS, and an understanding of what teachers knew about the DIBELS has been reported. The next theme that was identified for reporting was focused on how teachers are using the data, whether DIBELS or otherwise, to make decisions in the classroom.

Of the data collected in the present study through teacher interviews, the topic of using data in the classroom can be discussed in terms of driving instruction in the classroom, identifying students who are in need of additional supports, deciding what additional supports to give to students who are struggling, and using data to determine placement into ESE programs or retention within a current grade. Generally, the teachers in the present study reported meeting with their Reading Coach at least three times a year – corresponding with the timing of the DIBELS benchmark cycles – in order to review the data and identify areas for the grade level team to focus their efforts.

Of the teachers who reported mostly positive perceptions and value on the use of DIBELS, all reported a preference for using multiple sources of data to inform or guide their instruction. Some of these teachers reported using DIBELS data to confirm findings from other assessments in the classroom. When DIBELS data were found to be different from the district or classroom assessments, these teachers reported efforts to understand why.

Of the teachers who voiced mostly a negative perception or value on the use of DIBELS, all reported a preference for using their district or classroom assessments to make decisions. The reasons for de-emphasizing the use of DIBELS for informing their instructional decision-making were varied. Some teachers cited delays in accessing their classroom DIBELS scores; by the time they received them, the class had already moved forward in the curriculum. Some teachers reported no value for DIBELS to inform instruction citing a perceived lack of reliability and validity of the DIBELS since it was a “timed test” or because it included subtest measures that “do not apply” to their grade

level. And finally, some of the teachers who explicitly did not use the DIBELS to guide instruction cited a lack of value in the information it provided beyond the information already obtained through their district or classroom assessments.

The pattern of responses with respect to valuing the use of DIBELS to make instructional decisions appeared to match the pattern found in which teachers valued the use of DIBELS as influenced by their knowledge of the DIBELS. Thus, if a teacher reported limited knowledge of the test or misperceptions of the use of DIBELS, they also perceived little value in using it for making instructional decisions. Teachers who voiced greater knowledge of the DIBELS and/or perceptions that were consistent with the true parameters of the test also perceived higher value in using it to guide instructional efforts. It's important to note that none of the teachers who found benefit to using the DIBELS to guide instruction reported using the DIBELS as a single source of information. These teachers all reported benefits to using the DIBELS along with their district and classroom assessment to identify student needs and making decisions in the classroom.

The teachers who reported positive value in using DIBELS to guide their instruction specifically reported using DIBELS for organizing their instructional groups and adjusting these instructional groups throughout the year. In this context, teachers often discussed a differentiated instructional approach promoted through the *Reading First* grant. More specifically, these teachers highlighted the use of reading center activities, developed by researchers at FCRR, during their reading block time. These activities consisted of various reading activities grouped and organized by reading skill type and level of complexity. Teachers could access the activity plans through either the

FCRR website or through a large binder available from the school's Reading Coach.

Teachers who did not value using DIBELS to guide instruction did not report any use of DIBELS data to align or even develop their instructional groups. As one kindergarten teacher commented, "By the time we get our scores back, we've already determined our groups." These teachers described DIBELS as "just another test we have to give," or unhelpful because it doesn't reveal anything new beyond the district assessment data or observations made in the classroom.

Of the 14 teachers who were interviewed in the present study, eight of them explicitly identified using DIBELS at least to identify which students should have access to either Title 1 personnel/services and/or access to supplemental curriculum programs. Of these eight teachers, some of them relied on DIBELS to determine specific skills that a student needed help with and then assigned the student to work on those skills either with an assistant in the classroom or through a supplemental program. The other teachers did not indicate using the DIBELS data to identify specific skills on which to focus, but rather generally used the risk indicators of the DIBELS PMRN reports to select which students would access extra supports. The teachers who did use the DIBELS data in this fashion expressed concerns about how to interpret the data and a feeling of just trying to give the student access to everything they could give to them in hopes that something would work. The following programs were reported by teachers as being used for students who are identified as in need of extra help using the DIBELS data: Great Leaps, SRA Open Court, Head Sprout, teacher developed reading activities (cited by teachers

who used DIBELS data to identify specific skills to focus a student on), and *Reading First* activities (i.e., reading center activities binder).

Five kindergarten teachers reported specific use of DIBELS data – in some cases along with district assessments and classroom assessments – to make decision about whether a student would be retained, promoted to next grade level, and/or referred for psychoeducational evaluation to determine eligibility for special education. None of the first grade teachers reported any use of DIBELS at their school for making retention decisions or referring students for evaluation. One first grade teacher indicated only district assessment data were used at her school for deciding retention of students. Many teachers across both grade levels reported an expectation among student services staff (e.g., school psychologist, social worker, etc.) that DIBELS data be used for determining how students are progressing in response to interventions being provided to them in the classroom.

Within the theme of using data, teachers shared their views on the use of the Progress Monitoring and Reporting Network (PMRN) – a web-based program developed and maintained by the Florida Center for Reading Research which provides educators with graphs and reports for analyzing their DIBELS data. All teachers reported knowledge of this technology and indicated using the reports generated from it. However, the level of use found among teachers was substantially variable. All of the teachers reported using two specific reports the most: The “parent letter” and the Class Status Report (discussed later).

The parent letter is a report provided by the PMRN which involves a standard text template for communicating to parents what their student's score on the DIBELS was on the most recent cycle and how to interpret the results. It also provided parents brief ideas for how to support their child's learning to read at home. All teachers reported high value for the use of this specific report because it was easy and efficient in communicating to parents what the DIBELS scores mean. Some teachers, in order to make the report even more useful for parents, highlighted specific areas of the report to direct parent's attention to key information.

Few teachers indicated accessing the PMRN directly themselves in order to evaluate their class data. And among the few teachers who did access the PMRN independently, they reported accessing the PMRN mainly after each DIBELS assessment cycle. Most teachers indicated no need to access the PMRN because the reports they use from it are already provided to them by the Reading Coach (e.g., Class Status Report). And of the ones that expressed comfort and availability to print out their own reports, none of them chose to because they did not have access to color printers.

Because the PMRN reports utilize a color coding system of red, yellow, green, and blue to assist educators in scanning their data for analysis, the lack of access to color printing was found to be a barrier to teachers accessing the PMRN themselves. Of the few teachers who access the PMRN themselves, they indicated only knowing how to use the colors and scores to identify students in need of additional help. All teachers reported, regardless of their comfort level for using DIBELS or the PMRN, they would not be able to utilize the data as they do without the guidance and support of the Reading

Coach. All teachers expressed concern about the possibility of losing their Reading Coach when the grant expires, citing concerns about the impact it would have on their school.

Of the teachers who communicated a use of PMRN reports other than the parent letter, all indicated a high use of the Class Status Report. This report provided the scores for each of their students on the subtests given for a particular assessment cycle. Each student's subtest scores were color coded as described above and their overall performance profile across the subtests is reported as the "recommended level of instruction" (i.e., intensive, strategic, or initial). This report offered the option of organizing the data either by student name alphabetically, by instructional level needed, or by a specific subtest reported for that assessment cycle.

Some described a "box-and-whiskers" format (either a Student Grade Summary Report or a Class Grade Summary Report) that had been used in the past through the Reading Coach. All but one of those teachers that reported this type of format reported confusion about its style of data display and indicated little use for it. One teacher reported learning just recently from the Reading Coach on how to interpret that type of graphic display and further reported learning to focus on getting the boxes smaller and above the expectation line for a particular cycle as the goal. No other reports were mentioned by teachers.

#### *Comparison of Teachers and Experts on the use of PMRN reports – Case Study Review*

To answer the second research question, all teachers and two DIBELS experts were asked to review a case study. Experts were asked to review both a kindergarten and

first grade case study. Teachers were asked to review the case study appropriate for the grade level they teach. Results between teachers and experts were then examined for similarities and differences in the use of the DIBELS data presented through the PMRN reports used. Across all participants, teachers and experts, the common three themes that emerged in response to the questions asked through the case study review were: (1) using PMRN reports, (2) additional information needed, and (3) general comments about using DIBELS (experts only). Following the reporting of the findings for the first two of these themes a summary of additional comments offered by experts on the use of DIBELS data is presented.

*Expert review of kindergarten case study.* When presented with the three types of reports being used for the kindergarten case study, the experts utilized the cumulative report to evaluate the student's progress of skills and to evaluate the student's level of vocabulary. They reported the case study student was demonstrating a deficit in phonemic awareness skills as identified through the performance on the Initial Sounds Fluency subtest. Experts also indicated that because the case study reflected end of year performance, the data was useful for first grade teachers at the start of next year for incoming students prior to the first cycle of data collection for the next year. Using the Student Grade Summary Report, the experts demonstrated analysis of the student's performance in relation to the class and in relation to the expectation for that given assessment cycle.

Experts in the present study used the reports to identify student strengths and weaknesses. Across all reports, but specifically using the Class Status Report, experts

reported that the student had demonstrated proficiency in Letter Naming Fluency, required intensive help with phonemic awareness skills, and had demonstrated some knowledge of letter-sound correspondence. However, regarding letter-sound correspondence skills, both experts reported in interest in reviewing the student's protocols to identify patterns of errors on the NWF subtest. Further, both experts made general recommendations about providing the student with substantial interventions that are highly focused and narrow on the development of phonemic awareness skills.

Both experts reported a need or desire for additional information to better determine the student's needs or to better develop targeted interventions. Specifically, experts were interested in why the student was behind the peers in phonemic awareness skills. Was the student already receiving special education services? Was English a second language for this student? And if so, was he or she receiving ESOL support? Additionally, they asked how the student was behaving on the day of testing; is there any evidence that this information should be questioned for reliability? They also wanted to know what services or interventions have been attempted so far and how responsive the student was to those interventions. Finally, they asked if there available background information on the student? For example, did the student attend pre-school? Was the student born in the United States?

Regarding the cumulative report, experts wanted to know how the student went from 0 to 24 sounds on the NWF subtest. They asked if the student was receiving interventions specifically on letter-sound correspondence. They also questioned the reliability of the previous assessment cycle which reported the student had a score of zero

on the NWF measure. Finally, regarding the primary concern identified for this student, the experts wanted to know what phonological skills this student does have. They asked about what the student can do already? The experts reported the data on this student only indicates a concern about his/her ability to engage in phonological awareness tasks, but does not actually pinpoint where to focus interventions more specifically.

*Kindergarten teachers review of kindergarten case study.* When kindergarten teachers were asked to review the same case study, all seven teachers reported familiarity with the Class Status Report. Three of the seven teachers reported awareness of the Student Grade Summary Report. And only one kindergarten teacher recognized the cumulative report. All kindergarten teachers used the Class Status Report to identify student strengths and/or weaknesses. Responses by teachers were varied in most instances. However, all reported that the student had mastered his letter names. Some reported the student was “having trouble with sounds” based on the NWF score. Yet, others reported the student could perform well on the NWF, but that he/she “doesn’t know sounds” when referring to the PSF subtest. One teacher stated, “he knows his sight words” even though the ORF subtest was not provided to this student. A few kindergarten teachers highlighted the student’s need for help with beginning sounds and letter-sound relationships; similar to expert opinion. All kindergarten teachers provided suggestions to assist the student. Most of these were general statements of assigning a classroom assistant to work with the student one-on-one or in a small group. Two teachers indicated they would have referred this student for psychoeducational evaluation in consideration for special education placement.

Only four of the seven kindergarten teachers asked for additional information. Two of them wanted to know if the student's native language was English or if the student was receiving ESOL services. All four of the teachers wanted to know if the student was currently receiving special education services for either language or speech disabilities. One teacher asked for observations of the student during testing (e.g., any distractions?). One teacher asked if the student has had any history of short-term or long-term memory difficulties, or other "processing deficits." And one teacher wanted to eliminate questions about cognitive skills by knowing the student's IQ before assuming what needed to occur to support this student's learning needs.

*Expert review of first grade case study.* When provided with the first grade case study, as with the kindergarten case study, the PMRN reports were used to identify student strengths and needs. And as similar as with the kindergarten case study, the experts highlighted the information contained in the cumulative report to determine the student's progress through the year and examined the student's performance on vocabulary (PPVT) and comprehension (SAT-10). Specific to this case, the experts identified the student's primary area of need to be phonics, or letter-sound correspondence skills. Next, they targeted oral reading fluency as a need for intervention. The experts identified phonological awareness skills as a relative strength for the student. Based on the comparison of the SAT-10 and Peabody vocabulary results, and the student's ORF score, the experts hypothesized the student was compensating somewhat for comprehension despite having low vocabulary and low oral reading fluency skills. The experts viewed the case as requiring an individual focus as the Student Grade

Summary Report indicated the rest of the class was performing above the expected benchmarks on the NWF score. Finally, the experts highlighted the importance of looking at more than just the colors of the reports, but also the numbers in order to determine just how far ahead or below a student's performance was from the benchmark goal.

The experts wanted additional information as with the kindergarten case in order to determine what the student was already able to do before and to learn what interventions or services have been attempted to date. Specifically, regarding the students skills in phonics, the experts wanted to know what letter sounds the student did know. They wanted to view the student's NWF protocol to identify patterns of errors and blending skills. They wanted to explore more formally if the student's primary area of difficulty was a problem with decoding or a problem just with oral reading fluency as they indicated the student's NWF was just below the expectation for the assessment cycle. Experts advocated the student should be monitored more frequently for progress using DIBELS probes available from the official website for DIBELS from the University of Oregon. In this context, experts reported several resources were available on the FCRR website to assist teachers on the use of ongoing progress monitoring.

*First grade teachers review of first grade case study.* All first grade teachers recognized the Class Status Report and indicated using that report often. All reported knowledge of how to interpret the color coding system used in the reports. Reports were used by the teachers – the Class Status Report in particular – to identify student strengths and weaknesses. Only one of the seven teachers valued using the Student Grade

Summary report and demonstrated an understanding of how to use it. However, she reported that the box and whiskers format was mostly used to examine her whole class rather than a specific student in relation to the class. The report she was referring to was likely the Class Summary Report. All of the other six teachers reported no value in using the Student Grade Summary Report citing it was too confusing visually for interpretation. A few of these teachers recognized the cumulative report with some of them finding value in its use. One teacher in particular used that report first to look at the student's progress over the year; similar to how experts prioritized use of the same report in their examination of the data.

All of the first grade teachers used the reports to identify student strengths and weaknesses. Greater consistency was observed among first grade teachers' interpretation of the data as compared to kindergarten teachers. All of them identified the student's need in the area of phonics and reading fluency. They identified the student's strength to be in phonemic awareness. Most teachers described various activities to teach the student phonics, oral reading fluency, and sight word vocabulary. Some of these teachers indicated a preference for using alternative reports not included in the case study review in order to evaluate student progress: Class Historical Report and the Class Recommended Level of Instruction Report.

Regarding additional information needed or requested, five of the seven first grade teachers commented. Specifically, three first grade teachers wanted to view the student's NWF protocol to observe error patterns. Three of the first grade teachers wanted to know if the student was already receiving special education services. And if

so, was the student suffering from any disabilities that would impact their learning of phonics (i.e., ADHD, language impairments, etc.). Some wanted to know about the observations of the student during testing (e.g., any distractions?). Was this student learning English as a second language? Was he/she receiving ESOL services? Some wanted to view the student's error patterns on the Oral Reading Fluency subtest protocol. Finally, as observed with expert opinions, some of these teachers wanted to know what interventions had been attempted so far and what background information was available on this student.

*Expert opinions on the use of DIBELS data at Reading First schools.* Following the review of both case studies, each expert was asked to comment on particular comments made by them through the course of the case review in order to understand their perspective on what teachers should or should not be expected to do when using DIBELS data. From their view, DIBELS data did not tell one how to fix a problem; only that there was a problem and the area of reading development where the problem likely exists (e.g., phonics, or oral fluency). That is, because DIBELS was just a snapshot; a single moment in the student's life, it was not expected that the benchmark data be completely useful for instructional planning. However, the experts indicated that where a student is identified as having some difficulties in reading, it was important to examine that student's skills further; either through more formal diagnostic testing, or in response to interventions provided. Thus, it was the opinion of the experts that any student who was having difficulties as measured by the DIBELS, and there was confidence that the

data was reliable, ongoing progress monitoring should occur at a frequency appropriate for the level of need.

The experts reported that if the student’s recommended level of instruction was identified in the reports as “intensive” then the student should likely be monitored weekly to bi-weekly to monitor the effectiveness of interventions being provided to the student. They reported that if the student’s recommended level of instruction was identified as “strategic” then the progress monitoring should be occurring bi-weekly to monthly. It was the opinion of the experts that ongoing progress monitoring was more appropriate for instructional planning and decision making. Regarding ongoing progress monitoring, the experts were asked to comment on the number of samples needed for use as compared to benchmark assessment (i.e., ORF on benchmark assessments requires the use of three reading probes). These experts indicated that only one probe was needed per session for use when engaging in progress monitoring.

Experts also indicated that when analyzing the data through PMRN reports, it was important to determine if a student’s area of difficulty was isolated to just that student, or reflective of a larger number of students in the class or even the same grade level. This view was consistent with reports by teachers who meet with their Reading Coach as a grade level team to examine the DIBELS data across the grade level and classroom levels. They advocated educators to make hypotheses about why patterns of need exist and to then establish plans to improve outcomes; much like what some teachers reported was occurring when their grade level team meets with the Reading Coach.

In this same context, the experts advocated for educators to use DIBELS through a multiple data-source approach; using the data with existing data from other assessment to make instructional decisions. The experts also indicated high value in teacher's knowledge about a student's background for use when interpreting DIBELS information. However, these experts also expressed concerns regarding teachers' ability to make use of multiple sources of data, especially when these sources do not correspond initially; citing most teacher training programs do not adequately prepare teachers for how to analyze such data. Using DIBELS data only to make instructional decisions was cautioned against even though DIBELS was highly useful for making some broad conclusions about how to support a grade level, class, or student.

It was this concern about teacher's need for support to engage in more formal data analysis and instructional planning that the experts indicated awareness within their organization that much variability existed among schools in the ability to use data to guide instruction (not just with the DIBELS). They both reported that the Reading Coach was an invaluable source of support for schools and teachers to not only provide support on the interpretation of data, but also to help with using the data to make instructional decisions and to also provide ongoing professional development support for teachers. In relation to ongoing progress monitoring, the experts stated such activities were not happening nearly as much as they would have liked to see, but it was hoped, through ongoing support by their organization and from the schools' Reading Coaches that educators would be more prepared and influenced to engage in ongoing progress monitoring.

### *Attitudes and Perceptions Among Persons Other than Teachers*

The third research question was investigated through focus groups involving (1) Reading Coaches, and (2) “specialists.” Specialists in the present study involved a mix of student services personnel from the areas of school psychology, academic diagnostics, and ESOL. These specialists worked as itinerant staff members who were assigned as many as two to perhaps five different schools, with the exception of the ESOL teacher who was assigned to only one school. Their view of DIBELS was expected to provide a relative comparative view not only among different *Reading First* schools, but also between *Reading First* schools and non-*Reading First* schools.

Inclusion of this research question not only allowed for an examination of non-teacher perspectives on the use of DIBELS, but is also provided a way of examining the validity and reliability of teacher perceptions through a triangulation using multiple sources method (Miles & Huberman, 2002; Patton, 2002). Results of both focus groups were organized in the same manner as that used with teacher interviews for two reasons. First, it provides a consistent organizational framework. Second, the focus groups occurred after all teacher interviews had been completed; thus the focus groups were more effectively used for comparing non-teacher perspectives with those of teacher perspectives.

*Climate and culture of schools.* Reading Coaches and specialists both reported a belief that teachers were very overwhelmed in the district. Reading Coaches perceived teachers as not having enough time to complete their assessment responsibilities and having too much paperwork to complete. Reading Coaches perceived teachers viewing

the DIBELS as just one more thing imposed upon them. Reading Coaches perceived teachers having limited value for the DIBELS because it was one more thing they had to do which interfered with their instructional time.

Specialists characterized the climate as being very punitive for teachers which led to pressures to “teach to the test” in order to avoid negative professional judgments by their school administrators. Reading Coaches did not perceive a punitive environment, but cited barriers in the consistent use of DIBELS among teachers occurring from the requirement among teachers to engage in and collect data on students that has little instructional relevance. This perspective was evident among some kindergarten teachers who cited specifically the district’s entrance assessment as being lengthy and having little utility for instructional planning.

Specialists shared their comparative perspectives among *Reading First* schools and non-*Reading First* schools. Specifically, they indicated that the number of referrals for psychoeducational evaluation for consideration of special education eligibility was much lower at *Reading First* schools than at non-*Reading First* schools and have dropped in number at *Reading First* schools due to the implementation of the grant. Specialists also reported that teachers at non-*Reading First* schools were more likely to use DIBELS data to refer a student for special education consideration even when other assessment data suggested the student was performing at grade level.

Academic diagnosticians who participated in the focus group for specialists reported on patterns of student performance between those referred for evaluation from *Reading First* schools and non-*Reading First* schools as strikingly different. They

reported students at *Reading First* schools were not demonstrating as much difficulty in phonemic awareness and phonics skills as those referred at non-*Reading First* schools based on the nationally norm-referenced standard achievement tests they usually administer for students referred for psychoeducational evaluations. They attributed this observation to the increased amount of explicit and direct teaching of phonemic awareness and phonics skills at *Reading First* schools.

Both Reading Coaches and specialists emphasized the importance of school leadership setting the right climate and culture for using DIBELS data effectively. Reading Coaches in particular, being assigned to only one school, full-time, had the most to say about this topic. They stated a high value for data analysis was likely to exist, school-wide if the administrators for that school valued such data analysis and utilization along with holding staff accountable for using the information generated from such analyses. Reading Coaches reported the issue of why some leaders may not demonstrate explicit support for DIBELS had probably less to do with their perceived value as much as it was the competing demands of other responsibilities placed on them by the school district. That is, Reading Coaches perceived some principals were not able to demonstrate as much open and explicit support for the use of DIBELS (e.g., attend data analysis meetings with grade level teams) because they were burdened with much responsibility in other areas of school operation.

It was the opinion of Reading Coaches that in many cases, to help support school leaders in the use of DIBELS data, it was necessary to take the data to the principal and communicate often with them. Reading Coaches saw their role as important for this set

of efforts and felt an obligation to demonstrate, consistently, the utility of the DIBELS data. Further, they perceived some principals may not have to review the data often, but entrust a few to do that and provide input to staff. Overall though, Reading Coaches articulated a need for a school culture that generally embraced the use and analysis of data to make instructional and curricular changes. They insisted that teachers and administrators alike, many times, need to be led to using the data.

*Supports/Resources available to support use of DIBELS.* As observed consistently through teacher interviews, specialists identified the Reading Coach as the single most important variable supporting teacher's effective and efficient use of DIBELS data. Specialists cited Reading Coaches as having the responsibility for providing training on the use of DIBELS assessment team each year and in some cases before each benchmark cycle to ensure standardization procedures are known and followed for administering DIBELS.

A unique perspective found among specialists was that they perceived the Reading Coach should have had greater involvement and a stronger role in the coordination and development of interventions for students who were having difficulties in reading. Specifically, specialists cited the availability of a binder of student center reading activities provided by the Florida Center for Reading Research, but indicated a perception that teachers have very little time to utilize such a resource effectively enough. They suggested the Reading Coach could help with this by being familiar with the activities and offering recommendations to teachers when they meet with them to go over their grade level DIBELS data.

When asked to comment on their job description, Reading Coaches cited what amounted to a long list of responsibilities. These included the provision of technical support (e.g., data collection, analysis, and utilization) at the grade, classroom, and individual levels, modeling lessons involving research-based interventions or strategies, ensuring fidelity and integrity of instructional approaches or programs being implemented in a grade level or classroom, support for teachers who have ongoing questions about assessments or instruction, on-site training on the use of DIBELS, demonstration of the utility and value of DIBELS data, and coordinate data collection efforts. Regarding the demonstration of the utility and value of DIBELS data, Reading Coaches reported a variety of methods used to meet this job responsibility. Specifically, they cited meeting with teachers individually or in grade level teams. They engaged school leaders and leadership staff through meetings about the data. They help develop action plans to address student growth and improving outcomes. And they also cited use of the students' actual data protocols with staff in order to assist staff efforts to interpret the data.

Regarding the coordination of DIBELS data collection efforts, Reading Coaches confirmed the benchmark assessments as occurring three times a year. They also reported on efforts to engage teachers in ongoing progress monitoring activities. Reading Coaches reported teachers were generally starting to see the value of progress monitoring. Reading Coaches voiced uncertainty about how well teachers would embrace further progress monitoring efforts without support and encouragement to do so. When asked about their perceptions of why teachers may be reluctant to engage in more frequent

progress monitoring, Reading Coaches suggested it may be due to fear of results and/or due to the limited time available among all the other assessment responsibilities they have.

Reading Coaches cited efforts to increase teachers' value of and use of DIBELS. They reported one approach involved working with the teacher to understand a student's background or conditions of the testing situation in an attempt to understand why a student did not perform as expected and to develop interventions to support the student. Reading Coaches indicated that they believed teachers found value in data analyses when they saw a direct link to making plans for improving performance of students that involved others so that the teacher was not alone to carry out all aspects of the intervention. Reading Coaches also reported that it was beneficial to hold individual conferences with teachers and grade level teams to help them interpret the data and determine action plans in response to the data collected on students.

Specialists also voiced their perceptions on what they felt helped to increase teacher value of using DIBELS data. They indicated teachers were more likely to perceive value in using DIBELS if they perceived they had the time to do so. They also believed teachers were more likely to access the PMRN if the teachers had consistent access to color printers. Specialists reported that teachers appreciated the use of graphs and preferred to interpret the DIBELS data through graphical representation. However, because of the overwhelming amount of testing that was taking place in the schools, Specialists felt teachers appeared less motivated to engage in such analyses. Overall though, Specialists perceived teachers value and use of DIBELS to have increased with

each successive school-year; an perception that was also stated by some of the teacher who participated in the present study.

Specialists described several approaches that they found to be helpful in leading teachers to more acceptance and increased value for using DIBELS. One such strategy was to follow up with teachers after each assessment cycle to share the results quickly and explain any qualitative observations of the student that were written on the protocols. Another approach that was described was to sit with the teacher and review the protocols on students who were struggling and assist them in analyzing the patterns of data. Finally, one approach that had been observed to be successful among the Specialists in helping teachers to value the use of DIBELS involved going to the classroom before assessment cycles to give students advanced understanding of what they would be asked to do and to give directions for how to participate in the assessment process. This last approach was consistent with reports by teachers on the value of having DIBELS assessors visit the class to ensure the results obtained were valid and reliable.

*Teachers' knowledge of DIBELS.* Reading Coaches and specialists were asked to share their opinions on teachers' understanding of DIBELS. Specialists reported, similar to reports by teachers, that schools have evolved in their acceptance and use of DIBELS over time. Specialists and Reading Coaches confirmed previous teacher reports regarding teachers' concerns about students being tested by people other than the classroom teacher and about being timed on the DIBELS. Reading Coaches reported believing that teachers perceive the DIBELS as another high-stakes test that students need to pass. Both focus groups reported teachers were concerned about students being

timed and that teachers saw less validity in the use of DIBELS because of its timed aspect.

Regarding students being tested by persons other than the classroom teacher, Reading Coaches and specialists also indicated a concern about students being unfamiliar with the person who was testing them. Several procedures were shared that were used at schools to ensure students were demonstrating their true abilities during the DIBELS assessment. Reading Coaches reported spending time in the classrooms – especially kindergarten classrooms – prior to testing cycles in order to help students become familiar with who will be testing them and/or help them understand what they would be asked to do. Additionally, Reading Coaches spent time trying to help kindergarten and first grade teachers understand that the first cycle was not a reflection of their teaching as much as it was a measurement of the students’ incoming knowledge and ability to follow directions. Reading Coaches reported teachers seemed to feel pressured that they would be held accountable for the first DIBELS cycle despite the Reading Coaches’ attempts to alleviate such concerns.

Regarding student reactions to being testing, Reading Coaches and specialists reported very little concern beyond making sure students were familiar with who was testing them. Aside from that, no major concerns were reported by the focus groups about students being timed on the DIBELS. Rather, both groups reported students often appeared very comfortable and willing to engage in the task because they were used to the assessment. For example, specialists reported students in all grade levels often indicated they’ve seen the NWF “SIM and LUT page” by automatically reading “SIM”

and “LUT” before the instructions were given. That is, specialists and Reading Coaches did not report, as did some teachers, that students had any anxiety or concerns about being asked to read nonsense words.

Reading Coaches and specialists reported their perceptions of teacher’s use of the NWF subtest and their efforts to support teacher understanding about the NWF subtest. First, both groups reported that they thought teachers perceived the NWF subtest as confusing because teachers believed students were getting confused when trying to make real words out of the nonsense words as they’ve been taught to do with unknown words. However, both Reading Coaches and specialists reported students had little concern participating in the NWF subtest. They indicated that after they students got used to it, they no longer tried to read them as real sight words. Specialists emphasized the importance of the Reading Coach to help teachers and staff to understand the importance and usefulness of the NWF measure.

Reading Coaches reported trying to educate teachers about the NWF subtest being a measure of decoding skills rather than sight word reading. And though Reading Coaches understood what teachers were trying to express, Reading Coaches reported observing a more systemic pattern among students on the NWF subtest. Specifically, Reading Coaches reported seeing large numbers of first grade students decreasing in their NWF performance during the second DIBELS assessment cycle. According to Reading Coaches this finding was related to students being taught long vowel sounds at that time of the year – observations of students over-generalizing a new skill. Thus, when observing the errors among first grade students on the NWF subtest, large numbers of the

students were using long vowel sounds. Reading Coaches reported this phenomenon often had disappeared by the end of the year during the last DIBELS assessment cycle.

Both Reading Coaches and specialists agreed that teacher access to the scoring protocols was an invaluable method for identifying students instructional needs related to emerging phonics skills. Specialists and Reading Coaches advocated for DIBELS testers to take detailed notes about the students they've tested or some kind of indicator to remind them to go back to the teacher and share the information beyond the score. Both groups reported high value in the use of the student protocols to help make the data more valuable for teachers and to assist data analysis beyond the color coding system.

At non-*Reading First* schools, specialists reported observing teachers refusing to use the results of the NWF subtest because they believed teachers disagreed with using nonsense words to teach reading. Specialists further reported that it appeared teachers at non-*Reading First* schools had less understanding of the value and utility of the DIBELS. And aside from the use of nonsense words, Reading Coaches and specialists reported a unique concern in that the directions provided to students on the NWF measure offer the student the choice of either reading the nonsense words sound-by-sound, or as whole words. Some students reportedly will sound out the word sound-by-sound and then read the whole word; thus losing time and negatively impacting their final result. Both groups questioned whether such patterns are truly the emerging characteristics of the student's performance or if they merely demonstrated what was modeled for them in the practice trials.

Both Reading Coaches and specialists shared ideas on how they've attempted to help support teacher understanding of the DIBELS and increase value and utility for DIBELS. Both groups reported that sharing the research on the development and characteristics of the DIBELS would be, and in some cases had been, valuable for helping teachers to see the utility of DIBELS. Specialists believed it would offer teachers greater context and a "big picture" understanding for why the DIBELS was so widely encouraged for use. Specialists also indicated that school administrators need to know the research on DIBELS as well. Specialists believed that even though an awareness of the research would not help them know how best to interpret the assessment results, it would at least help them appreciate its use as a reading assessment.

Reading Coaches added another approach to helping teachers use DIBELS by working with them in the classroom to set up various center activities and model instructional approaches (e.g., Word Work). They also reported helping the teachers to use student assessment data to assign students to specific center activities. They reported working with teachers to analyze the NWF fluency data and other DIBELS subtests to help teachers find the patterns in the data.

*Collecting and using assessment data.* In general, Reading Coaches indicated that they thought teachers felt more comfortable and trusting of the DIBELS data when the same person collected the data each cycle for their class. Specialists identified different data collection procedures at different schools. The choice of which procedure was used at a given school appeared to have been adopted by the influence of what seemed possible at that time for the school (e.g., staff available, schedules, structural arrangement

of the building, etc.). Regardless of which procedure was used, specialists reported that the efficiency in the process of collecting DIBELS data had steadily improved year to year. Both groups reported the same two types of procedures for collecting data school-wide as reported by teachers.

Given their unique position in the school district, specialists described unique views on assessing special populations with the DIBELS (e.g., ESOL, special education, etc.). They reported that it was helpful if the same person who worked with the student or students continued to be the person who assessed them in the future – or at least made sure the students were very familiar with the assessor. They perceived students who had special needs or circumstances would show a truer or reliable performance with someone they knew and had worked with. Specialists reported it was helpful to have a system in place that communicated specific student circumstances or characteristics that would need to be considered before working with such students regardless of who collected the information. And additionally, specialists reported value in sharing results with teachers immediately after testing; including observations and qualitative notes.

The perceptions of Reading Coaches on the teachers' use of DIBELS data were more extensive. Overall, Reading Coaches agreed that teachers' acceptance of DIBELS had evolved over time with each successive year of implementation. In the beginning, according to this group, teachers accepted DIBELS data that validated their expectations of a student and did not value DIBELS data for students who did not perform well on it. Reading Coaches reported it took approximately 2-3 years for teachers to see the value of DIBELS and they observed some teachers found the DIBELS more reliable because

someone else conducted the assessment which was similar to that reported by one of the teachers interviewed. At the time of the focus group, Reading Coaches reported teachers were just starting to show an interest in learning how to administer the DIBELS in order to see for themselves how students perform and respond to the assessment.

Reading Coaches reported that they believed teachers were so inundated with data that they were unsure about how to use all of it together. They suggested that much variability existed in teachers being willing to take the next step in data analysis and data utilization; an observation that was consistent with the DIBELS experts who were interviewed in this study. Reading Coaches advocated more support and training be given to teachers on how to use multiple sources of data to make decisions about a student's needs. They reported that many teachers were still having difficulty seeing the correlation between fluency and comprehension; not to mention how to interpret the DIBELS scores.

Reading Coaches indicated some teachers were putting too much emphasis on the DIBELS at the expense of ignoring other measures/assessments. This observation was in contrast to teacher reports about feeling the primary focus on the DIBELS emerging from administrators and student services personnel. Specialists shared similar statements for teachers having access to more training about the DIBELS. They believed that teachers showed a preference for Running Record assessments in first grade. They thought teachers did not have enough information about the DIBELS to appreciate its use as a reading assessment. They also reported a belief that teachers needed more assistance on

how to use DIBELS data to organize their instructional groups and to modify these student groups through the year based on the data.

Reading Coaches shared their views on the use of PMRN reports by teachers and administrators. Overall, this group reported believing that teachers needed continued help and training on how to use the PMRN reports. It was the perceptions of Reading Coaches that KG and 1<sup>st</sup> grade teachers often appeared more proficient at using the reports. This group indicated some teachers were simply more proactive in using them and is more independent at accessing their own reports online.

Reading Coaches thought the increasing proficiency and use of PMRN reports was related to the evolving acceptance and understanding of DIBELS among teachers as a whole. However, they reported the skills involved in data analysis were less tangible; much like the observations reported by DIBELS experts on this topic. Reading Coaches advocated for more teacher supports to teach staff how to analyze data at a level of independence. In summary of the PMRN reports, Reading Coaches stated that they believed teachers would embrace using these reports if they saw the value in them. Reading Coaches expressed concern about what supports would exist for teachers on how to find value in the reports when the grant expired.

Specialists offered additional recommendations on how to support teachers in their development of data analysis skills. They suggested teachers needed to put more emphasis on student growth rather than focusing solely on student level or color of performance. Specialists argued educators needed to be watching the student's trend rather than a single performance on one assessment cycle. Reading Coaches also stated a

need to focus on student trend and growth rather than a single snapshot in time. Thus, specialists and Reading Coaches saw a need for teachers and schools as a whole to engage in more ongoing progress monitoring and attend training on how to utilize multiple sources of data. These observations and recommendations were consistent with those offered by the DIBELS experts who participated in the study.

Regarding ongoing progress monitoring, specialists reported that they believed teachers were starting to engage in some progress monitoring, but that teachers were having problems with how to do it for a variety of reasons. Specialists described teachers as having limited time available to engage in progress monitoring. Specialists perceived teachers weren't getting out of it as much as when someone else had done it and brought the information and observation notes to the teacher. Specialists also cited the lack of available space to conduct the ongoing assessments without distraction or interruptions. They also indicated a lack of personnel to assist the teacher in covering classrooms while they engaged in progress monitoring activities. Overall though, specialists perceived teachers at *Reading First* schools seemed more ready to understand the role of progress monitoring than teachers at non-*Reading First* schools.

A final area of discussion regarding teacher use of DIBELS data involved cautions and advice from specialists and Reading Coaches, respectively. Specialists, when asked to comment on their perspectives on teachers' administering the DIBELS themselves, gave a one-word response: "unrealistic." Specialists reported teachers would need extra personnel to watch the rest of the class or else the reliability or validity of the DIBELS may be compromised due to breaks in standardization or assessment error. This

group cited the current state of testing for kindergarten teachers using the district assessment which required one-on-one testing as a barrier to having kindergarten teachers take responsibility for administering the DIBELS. Teachers have already reported – and specialists gave similar observations – that because the district assessment in kindergarten took several weeks to complete they were challenged to both test and conduct instruction at the same time. The specialists argued that even if kids were provided with independent activities, a teacher conducting assessments alone using DIBELS could take weeks to complete a whole class.

However, specialists indicated some kindergarten and first grade teachers who have taken the initiative to conduct their own DIBELS assessments have been observed to not need someone to come back and interpret the data for them. The overall concern or caution offered by the specialists was that a standard requirement for all teachers to administer the DIBELS without providing them support or at least removing some of the current assessment responsibilities was unrealistic; at the very least because of the large variability that currently existed among teachers abilities to conduct the assessment themselves. The specialists suggested teachers should be allowed to volunteer rather than be mandated to conduct the DIBELS assessment on their own following the expiration of the grant.

Reading Coaches gave several pieces of advice and made cautionary comments on the current and future use of DIBELS at schools. They reported critical decisions were being made on very small snapshots of student performance such as retention and special education consideration. Reading Coaches reported concerns about the fidelity of

use among non-*Reading First* schools. Reading Coaches also expressed concerns about teachers administering the DIBELS themselves for the same reasons as cited by specialists; limited time and little assistance. On the topic of PMRN reports, Reading Coaches advised the PMRN should generate graphs that reflect the correlation between students' oral reading fluency and later performance on 3<sup>rd</sup> grade FCAT to influence teachers' use of DIBELS.

In summarizing these concerns and advice for current and future use of DIBELS among kindergarten and first grade teachers, Reading Coaches emphasized their role's importance to make sure new teachers have the support they need to learn how to use the DIBELS effectively. They cited the need to support teachers who are at different levels of understanding and proficiency in the use of DIBELS. Reading Coaches were concerned about who would provide guidance and direction for coordinating data collection and analysis activities when their role was discontinued. They reported that currently at their schools there was no one available or capable to adopt these responsibilities – “it's a full time job,” as one Reading Coach added.

If teachers were asked to take over the responsibility to administer the DIBELS, Reading Coaches stated teachers would need more planning time, other assessment requirements be taken away and/or additional personnel to help with covering classrooms during assessment cycles. Reading Coaches argued the district needed to prioritize the assessments they are demanding teachers to use.

### *Analysis of Hypotheses/Researcher Expectations*

At the start of this study, the researcher provided several hypotheses about what would be expected to be occurring in the field prior to conducting the present study. These hypotheses were provided for guiding investigative efforts and adding credibility to the study, as well as to guide efforts to minimize any potentially negative influence of the researcher's biases. The hypotheses included the following:

1. There exists substantial variability among teachers' perceived value of DIBELS in assisting their students' learning needs.
2. There exists a relatively high level of variability in the perceptions of non-teacher participants involved in the implementation of procedures for using DIBELS at the school-building level.
3. Given the various other assessment tools used at the classroom level and the overlapping schedules of providing those other assessments (e.g., FCAT, county-wide assessments, Lexile assessments, Running Record assessments, etc.) within a school district, teachers' perceptions of using DIBELS are negatively impacted.
4. Teachers understand what DIBELS is and what it measures, but are discouraged or unsure about how to best utilize the data obtained.
5. Given the multiple competing demands and seemingly fast paced nature of school activities, teachers are not accessing their class/student reports on the PMRN, but rather are provided such reports, if any, from the school's Reading Coach or school-based DIBELS team.

6. A low level of direct involvement and use of PMRN reports serves as a barrier for utilizing the data effectively or efficiently.

Having provided the results of the study in relation to the stated research questions posed from the start, the above hypotheses were analyzed. Evidence was found for confirming the first hypothesis through teacher interviews. However, variability was found not only across teachers but within each teacher participant. It was not possible to simply classify each teacher as merely being in favor for or against the use of DIBELS overall. In fact, all teachers voiced various reasons for positively valuing the DIBELS, and concerns for its use.

Hypothesis number two cannot be confirmed given the results that were obtained through Reading Coach and specialist focus groups. Review of transcripts from those focus groups actually found very little, if any, variability in the perceptions of teachers use and value on the DIBELS.

Results from teacher interviews would suggest that hypothesis number three is credible. Several teachers spoke of the pressures and lack of time related to the numerous assessments that were conducted in their schools or classrooms. In particular, kindergarten teachers reported that their district assessments took as much as three weeks to complete a whole class. Some evidence existed that teachers were not necessarily against the use of DIBELS, but rather, because of the various other assessment demands and responsibilities imposed by the school district, they perceived the DIBELS as just another test they have to use. Some teachers reported they discovered how to use DIBELS with other assessments while other teachers continued to struggle with how to

make sense of data from the DIBELS when it does not correspond with other assessments they use.

The fourth hypothesis can neither be confirmed nor disconfirmed. Rather, it seemed that teachers who had a relatively strong understanding of the DIBELS valued the use of PMRN reports more than teachers who had less understanding of the DIBELS. However, it appeared all teachers were having difficulty linking the results of DIBELS with instructional decision making. Moderate variability was found in the ability to use the DIBELS data to make decisions, but all teachers interviewed in this study indicated a need for more support towards this objective. Additionally, some teachers asked for more support through available school staff to analyze the data for them, while others asked for more training to learn to analyze the data themselves.

Some evidence existed to support the fifth hypothesis. All teachers reported limitations to using the reports available to them through PMRN. Some of those limits were access to color printers to make optimal use of the color code system the PMRN reports use for visual analysis. Given this, teachers were reported to be highly dependent on the Reading Coach or other school personnel for accessing color reports. Other limits involved teachers forgetting their password and/or usernames. Some teachers indicated not having enough time to look through the website and make use of the information that was available to them. It appeared that those teachers who had a higher value and understanding for the DIBELS reported accessing the PMRN more than those who valued the DIBELS less. All of the teachers indicated accessing the PMRN to use the parent letter which was reported to be very valuable to all the teachers for communicating

to parents what the DIBELS measures and how parents should interpret their child's scores.

Regarding the last hypothesis stated at the start of the study, it cannot be confirmed or disconfirmed. Rather, it appeared several variables existed as barriers to teacher's use of DIBELS data. Though some teachers reported having difficulty making the most of the PMRN/FCRR website, many of them indicated a lack of training on what reports were available and which ones were suitable for different types of analyses.

Aside from the PMRN reports, it seemed evident that teachers preferred their own district or classroom assessments over the DIBELS because they administered them and it allowed them to observe the child in the testing session. Teachers voiced positive value in being able to observe the student during testing in order to gain an understanding of what the student can do. But most of them declined to be trained on how to give the DIBELS citing time and workload as barriers. Even though teachers described the district assessments to be burdensome, time consuming, and/or an interruption of instruction (all of them indicated the DIBELS was more time efficient) many still preferred to give the district assessment over the DIBELS because they were the ones that gave it. As was observed in statements made by Reading Coaches and specialists, it may not be realistic to expect teachers to take responsibility of giving the DIBELS unless something else is removed from their responsibility regarding assessments.

#### *Analysis of Unanticipated Topics*

Several topics emerged throughout this study which were unanticipated by the researcher. The first concerns the use of the NWF subtest. It was surprising that even

among teachers who reported high value for the DIBELS had also reported concerns with using the NWF subtests. Teachers reported that students were confused about how to perform on this test because they were attempting to apply strategies they were taught in the classroom for identifying unknown words. Some of these strategies involved using picture clues or context clues. It was also reported that students were generally taught strategies to sound out unknown words and think of words they knew that were similar and/or made meaningful sense.

It was interesting that Reading Coaches, though they did not share the same heightened concern for the NWF subtest, indicated a larger pattern among first grade students when they were being taught long vowel sounds in the classroom and then try to apply long vowel sounds during the NWF subtest. The Reading Coaches indicated this appears to be most evident during the second DIBELS cycle in the first grade, but then was less of a concern by the end of the year.

Another unanticipated concern was teachers' perceptions of the DIBELS as a "timed test." Only a few of the teachers characterized the DIBELS as a fluency measure and were able to compare and contrast it with other assessments as a difference between measuring accuracy and fluency. But most teachers perceived the timing of students either unnecessary at the least, or developmentally inappropriate at the extreme. Some teachers expressed great concern about students being anxious or stressed because of being timed on the test. And yet, no teacher could recall a specific instance where a student voiced fear or concern about being tested in the DIBELS.

Reading Coaches and specialists did not report any instance of students feeling anxious or intimidated by being tested with DIBELS. However they did report that for the first cycle in kindergarten, students were not familiar with it – and teachers were advised that the results for kindergarten students on the first cycle was not a measure of their classroom as much as it was a measure of what students skills were upon entering school and how well they could follow directions. By the second cycle, there was no concern reported by Reading Coaches and specialists on the testing of students with DIBELS. When asked why teachers perceived the timing aspects of the DIBELS negatively, Reading Coaches and specialists suggested teachers often perceived the DIBELS as a high stakes test that students were supposed to pass.

Teacher use of the parent letter was unanticipated. All teachers reported using it and valued it as a tool for communicating with parents about their child’s reading skills. Some teachers reported modifying the reports to highlight the sections most relevant for parents. Others reported value in the recommendations it provided parents to support their child at home. And others reported value with respect to the parent letter for how it described what the DIBELS are and what they measure.

It was anticipated that teachers who had been teaching at a *Reading First* school for at least two years of implementation would have had sufficient training on how to not only understand what DIBELS measures, but also how to interpret the scores. Though some teachers, as reflected through the case studies, interpreted the data in similar ways as the DIBELS experts, some teachers demonstrated a continued need for support. Mostly, it was reported that teachers were still very highly dependent on their Reading

Coaches for understanding what the DIBELS can offer them and how to use the data it provided. Many of the specialists reported concerns for the future use of DIBELS when Reading Coaches become unavailable following the expiration of the grant. A few teachers indicated they would likely be okay with interpreting and using the data without their Reading Coach but did not see how the school would be able to support the data collection efforts without the Reading Coach.

A final unanticipated result observed concerned the presence of unreliable DIBELS data. Reading Coaches and specialists both indicated a need to review standardization procedures among their data collection team to maintain a reliable collection of the data. Only one teacher of those interviewed indicated retesting students as an option when measurement error was suspected. DIBELS scores are more stable indicators of performance when repeated assessments are given (Kaminski & Good, 1998). The benchmark assessments serve conceptually as screenings for identifying students who may be at-risk for reading difficulties. When a student was identified as having difficulty and that identification was found to be in contrast with what was known about the student either by observation or other assessments, there did not appear to be any process in place at the schools sampled for checking the accuracy of data for that particular student.

Related to this issue was the use of DIBELS as a progress monitoring tool. DIBELS experts interviewed in the present study confirmed that very little progress monitoring was occurring, not only with DIBELS, but with other data sources in the schools. And yet, the DIBELS experts reported that progress monitoring data was most

useful for making instructional decisions for students rather than the benchmark data which only represents performance at one point in time. This perspective was consistent with the concerns voiced by Reading Coaches regarding high stakes decisions such as retention and/or access to special education based on DIBELS benchmark data.

Importantly, Reading Coaches, DIBELS experts, and specialists reported that progress monitoring with DIBELS was only just being introduced to teachers. Thus, it was not surprising that little progress monitoring was occurring. However, these non-teaching professionals expressed concerns about the appropriateness and/or possibility of asking teachers to conduct their own progress monitoring assessments with DIBELS when they were already engaged in a high level of assessments in their classrooms.

#### *Analysis of Qualitative Research Process – Researcher’s Reflections*

*General introduction.* This next section provides the reader an account of the researcher’s reflections through the course of the study as a means to offer an audit trail regarding the changing dynamics of the research process as experienced by the researcher. To provide a fluid reflection of the research process, this next section was written in the first person and written in chronological order. It is expected that the following will further provide the reader with a rich and thick description of the research study towards an evaluation of the credibility and integrity of the data and its interpretation (Patton, 2002).

*Research proposal and preparation for conducting the study.* The research proposal was presented in May of 2006. Several recommendations were provided by the doctoral committee to revise specific aspects of the study and its methodologies. These

revisions were completed by the end of the summer of 2006 with the exception of one recommendation which involved the development of more knowledge on qualitative research methodology. To meet this need, several months were used to further review the research literature on qualitative methodologies. Additionally, the field instruments that were developed for the study (e.g., guided interviews, focus group guides) were shared with researchers at the Louis De La Parte Florida Mental Health Institute (FMHI) who were recommended for their expertise in qualitative research design.

During the months of January and February of 2007, I consulted with research experts at FMHI to develop sufficient knowledge of qualitative methods and to ensure that my instruments were appropriate for use. Having received feedback from these expert researchers, I revised my instruments as needed and kept notes of all advice given for later consideration. One particular recommendation provided by these experts was in regards to the focus groups. They suggested that I enlist the help of someone trained in facilitating focus groups to provide greater credibility. Specific names were obtained for contact to solicit their support in addition to recommendations to contact the department of anthropology for possible help. I was unable to find anyone willing to help without requiring a modest fee for their service. Given the financial limits of the study, I was unable to hire anyone to facilitate the focus groups.

I shared this experience with a couple of doctoral committee members and I was advised that there was no expectation that I must have someone else conduct the focus groups, though it would be allowed if I could find someone. Having no success in finding someone with sufficient experience at an affordable cost, I prepared to conduct

the focus groups myself and enlist the help of graduate students who had experience in helping to facilitate as a scribe – to assist in keeping notes through the focus group meeting. Graduate students were identified with help from committee members and enlisted to participate at a later time when the focus groups were scheduled.

*Accessing the research site and participants.* Having received approval from the doctoral committee and while soliciting consultation advice from additional research experts, I engaged in the process of gaining approval from both the University IRB and the school district’s own research and accountability office. Though my original plans for accessing the schools and participants were approved by the University IRB, they were not acceptable to the school district’s research office. I was advised by the school district personnel that I would have to modify the procedures for selecting participants as I would not be permitted to contact teachers directly; though I could send invitations or similar documentation announcing the opportunity for participation. Further, I was advised by participating school district personnel that I would have to submit any documents for approval by the school building principal for authorization before obtaining consent from any teaching staff.

All of the above required changes to my proposal regarding procedures for recruiting participants and accessing sites. However, it also required adjustments to the manner in which I sought to purposefully include variability based on school performance as measured by schools’ Annual Yearly Progress (AYP) score – a score determined by performances on the FCAT. This information was presented to my

committee and I was permitted to revise my procedures for recruiting participants. The specific procedures were listed in the methods section of the present document.

Originally I had intended to randomly sample teachers from specific schools that would have been selected based on outcome scores on FCAT (i.e., AYP percentages). However, because of the constraints imposed by the school district, I was required to send invitations/announcements to all *Reading First* schools' principals for their approval and then ask them to forward the remaining materials to kindergarten and first grade teachers, their Reading Coach, and their DIBELS data collection team.

In addition, the school district required me to have a district sponsor for conducting the study. I was able to obtain support from one of the school district's supervisors of the *Reading First* program. Having shared with her the scope, purpose and intent of the project I then shared with her the constraints regarding participant recruitment. Given the large size of the school district and the number of schools that were potentially eligible for inclusion in the study, we decided it might be helpful, prior to mailing the invitation packets to all schools, that she contact principals at those schools first to demonstrate the school district's support for the study and secondly to announce in advance the materials that would be forthcoming. I found this assistance very helpful as I was concerned initially about principals receiving, unexpectedly, a large envelope of materials to review when their time was of course very valuable. It was also helpful because the list of potentially inclusive *Reading First* schools totaled 54 elementary schools. Thus, it was helpful to have a means of communicating in advance with so many schools at once to quickly and efficiently communicate approval and existence for

the study. I gave each school two weeks to review and consider the proposed study before following up with phone calls or emails to the school's principal.

At this point in time, FCAT testing was occurring and continued during most of March. I had anticipated minimal response in light of this context. Following the two-week window for review of the proposal, I emailed all principals along with providing a carbon copy to my district sponsor asking if they had received the materials and if they had an opportunity to consider the requests for staff participation. Very few principals responded, and of those that did, a few of them indicated they would not be able to participate, citing concerns with the timing of the year (i.e., nearing the end of the school year). Some principals did send an email response indicating they had passed along the materials to the specific persons identified in the cover letter – most of them chose to pass the information along to the Reading Coach to manage.

While waiting for potential participants to contact me, I organized my materials and prepared for data collection. I obtained a digital audio recorder and immediately began to see the benefit of such a device as the audio files could be uploaded onto my personal computer for ease and storage of the data. I also printed enough copies of consent forms, organized them along with copies of interview field notes, and a personal journal for use to record personal reflections of the study. During data collection efforts, I soon found the audio recorder to be more efficient and effective for recording my reflections and hunches. By recording my thoughts on the audio recorder, I was able to go back to them as needed and listen for emerging ideas and questions about the study.

Approximately one month passed before receiving the first emails and phone calls from teachers who had an interest to participate. As each teacher contacted me to participate, I made arrangements immediately to meet them at their school on a day and time of their choice. Data collection of teacher interviews occurred between the second week of April and the third week of May. During this same time, a few Reading Coaches and specialists contacted me with an interest in participating in the focus groups. Only one principal had contacted me by email to indicate an interest in participating in the focus group for principals.

Several Reading Coaches asked to participate and no concerns were noted in recruiting their participation. However, attempts to find a time and place convenient to all of them proved to be especially challenging. Given this, I contacted my district sponsor for assistance in finding a solution. She indicated a Reading Coaches meeting was scheduled for later in the month of May and that I could discuss with the potential participants their interest to participate prior to their staff meeting. This proved to be very effective as it already determined the place where the focus group would be held and all were willing to participate prior to their staff meeting. The Reading Coaches focus group was held on May 14, 2007.

Recruiting DIBELS experts also proved to be a relatively easy process. I contacted a representative of the state agency that provides training and technical assistance to schools on the use of DIBELS and presented the scope, purpose, and intentions of the study. I gave this person my contact information along with a request for help to identify and recruit two persons to serve as DIBELS experts. I was then

contacted by email within a week by two individuals. Interviews with both of these individuals were held via telephone, respectively, and recorded using the audio recorder for later transcription. Because the interviews with the experts were held approximately mid-way into teacher interviews, and though not explicitly planned from the start of the study, I did find it fortuitous to talk with each expert following the interviews about the information I had obtained thus far from teachers and found the experts very gracious with their time in sharing their thoughts about the findings thus far. These additional comments and views are listed in the results section as general expert opinions on the use of DIBELS.

Towards the last month of school for the district (May) I received many more requests to participate from teachers. However at that time, I was still concerned about the show of interest among specialists and principals for participation in focus groups. By the first week of May, I had received interest from five potential specialists – four of which were all from the same school who were willing to host the focus group at their school. To ensure a sufficient number of participants as well as a balance of different perspectives from different schools, I contacted several school psychologists and academic diagnosticians to solicit their participation in the specialist’s focus group. Of those that contacted me with an interest in participating, a few had met the criteria of being involved in a school’s DIBELS data collection efforts for at least two years. Following this round of recruiting, I had at that time, a pool of 12 participants for the specialists’ focus group. However, when attempts were made to find a place, date and time that would be convenient to all, five had to decline participation as they could not

travel away from their school campus and one had to decline due to personal events. The focus group for specialists was held on May 17, 2007.

By the first week of May, I had not received any requests to participate from principals, except for the one early in April, despite several emails. I had considered towards the end of April in revising procedures to try to engage principals in a one-on-one interview at a time and place convenient to them. However, prior to initiating this adjustment to procedures, I contacted seven principals by phone to determine the viability of this approach. Of the seven, three did not return my phone call, and the other four declined to participate, citing time constraints. As a result of this experience, further attempts to involve principals were suspended. Attempts to include them through the data analysis procedures (e.g., Member Checks) are described below. Site entry into the schools was made available through support from the district sponsor. However, her role in the district was parallel to principals and therefore likely did not offer any opportunity to ensure principal participation.

*Data collection and data analysis.* Going into the data collection process, I reminded myself of my own background experiences in using the DIBELS and my training as a school psychologist. Indeed, both of these sets of experiences had led me to highly value the DIBELS as a tool for helping to improve the reading outcomes of students and assist teachers in making instructional decisions. I also reviewed my hypotheses of what I expected to find occurring in the field. Because of my high value for the use of DIBELS, I wanted to make sure that I did not engage teachers in a debate

about how they should value it or limit myself to only hearing what I wanted to hear. This of course required much self-discipline and openness to other views.

So, in preparation for such openness, I reflected that although I perceived DIBELS is a useful tool, it was teacher's perceptions that were most important for understanding since they provide the instruction to students. I felt that I needed to have an appreciation for what teachers go through to best prepare myself to listen openly to their perceptions. So, I talked with a couple family members who are themselves teachers and talked openly about their experiences working as teachers and the struggles they encounter trying to meet their students' instructional needs. They expressed to me the pressures they feel in their jobs and the negativity that is sometimes presented in the public media about the quality of schools. They shared with me their concerns about how the schools are changing and how it makes them feel less valued as professionals. And of most interest was how they indicated to me knowing what to do to help students, but often feeling constrained to do so because decisions for how the school system should operate were made by those outside of the classrooms.

I valued these conversations I had with my family members because it helped me to avoid perceiving teachers as the source of blame for why they may not be using DIBELS or valuing DIBELS. Instead I was able to walk into each interview with a perspective that every teacher cares about their students and each teacher works very hard to meet their students' needs. These talks also served to emphasize a need to pay particular attention to how the systemic variables either supported or hindered teachers in their use of DIBELS. I walked into each interview assuming that every teacher values

data. I also was able to appreciate all the pressures they are under for student academic outcomes and imagine how it might make me feel to be under such pressure. I found myself more capable to see how competing demands and/or conflicting educational policies and practices could interfere with a teacher's selection of what tools they chose to use in the classroom. Further, to make sure that I maintained an openness and empathy for what teachers are experiencing in their profession, I wrote down these insights I gained and kept them in my possession throughout the whole study to remind me of my need to be open to other possible reasons for what may be happening in the field.

To further guard against the possibility of any possible bias on my part, I started each interview with a brief, but openly honest expression of the purpose of the study. All participants were aware of my employment in the school district as a school psychologist and were further made aware of my value for the DIBELS as an assessment tool. However, I also shared with each participant or group that although I had a favorable outlook on the DIBELS, I was not a teacher and therefore wanted to understand teachers' perceptions by speaking with them directly. I encouraged them all to be as honest as they could to share their own views, especially if they did not value the DIBELS, or any aspect of it. I believe that by having been open about my biases and role I was able to maintain openness to what teachers wanted me to know from their perspective.

Further, in accordance with my understanding of qualitative methods, and advice received from research experts on qualitative methods, I had planned a priori to maintain a consistent search for alternative views or themes, as well as divergent patterns and rival explanations (Patton, 2002). I had assumed going into the study that it would be possible

to identify teachers as either having a positive or negative value for DIBELS. However, as data collection proceeded, even from the first interview, it became increasingly clear that no one teacher could be classified on a simple dichotomous scale. This experience further reinforced the need to seek out alternative views, opinions, and more specifically limits to one's positive or negative view of the DIBELS.

During all interviews and focus groups, when an individual or group shared positive or negative views on the use of DIBELS, explicit questions were used to test alternative views. For example, if a teacher voiced positive value in the use of DIBELS and did not provide alternative views, then I would ask her explicitly to identify any concerns or negative values about the DIBELS. This consistent approach to seeking alternative views both within and across participant sources proved to be very useful as it helped me to see how complex participant views were on the topic of DIBELS. As mentioned previously, I found early on and throughout the data collection phase that perceptions of the DIBELS could not be classified as a simple like or dislike. Rather, their values and perspectives were related to variables operating in the schools. Some of these variables were anticipated and others emerged through the course of the interviews/focus groups.

Following each interview/focus group, attempts were made to try to transcribe the audio recording. However, due to the rapid time frame in which interviews and focus groups were conducted, combined with limited resources to support the transcription process, I found it very difficult to transcribe each interview/focus group before engaging in the next interview. I had investigated the option of hiring someone to transcribe

interviews/focus groups for me, but as was found with trying to hire someone to conduct focus groups, the fee for such services was prohibitive. So, to maintain a fresh perspective on the evolving aspects of the study, summaries of interviews were written and my reflections of them were recorded. This way, before entering into another interview, I was able to review the notes from the previous interview(s) and reflections in order to (1) listen for possible recurring themes, (2) divergent and convergent views, and (3) test the limits of specific unanticipated themes.

The use of a triangulation across sources method was planned a priori for the present study and discussed more specifically in the context of analyzing data later in the present paper. However, no explicit plans were made about the order in which interviews and focus groups would be scheduled as it was anticipated that several factors beyond my control would influence how they would be scheduled. It was fortuitous though that the focus groups occurred after the majority of teacher interviews had been completed (i.e., 11 by the time of the Reading Coaches focus group and 13 by the time of the specialist focus group). The timing of the focus groups was fortuitous because I was able to seek convergence and divergence on specific anticipated and unanticipated themes that arose during the teacher interviews.

On some very specific unanticipated themes brought up by teachers, I had expected the Reading Coaches and specialists to have different views from each other. However, I was surprised to find very little divergence of views on all themes being monitored between these two focus groups. By being able to seek non-teacher opinions on the use of DIBELS from two different groups of people after teacher interviews had

occurred, I found myself able to share specific quotes or comments provided by some teachers to check the validity of such statements among the focus group participants, as well as solicit their opinions about such statements.

Data analysis occurred concurrently with data collection procedures. I tracked anticipated and unanticipated themes by summarizing and reflecting on each interview. Early stages of analysis involved looking for recurring patterns of responses and identifying limits to the views or opinions shared by participants. Analysis in the early stages of data collection also involved keeping a pulse on my own organization, objectivity, and reflections as I tried to document my changing views and focus through the course of the study.

Throughout the study, I found it deceptively easy at times to not write something down or mention it into the audio recorder because many times my thoughts were so subtle or seemingly so benign at first that they did not register as “relevant.” When I found myself failing to record such thoughts, I would immediately write down or audio record my thoughts. Some of the ease in avoiding the recording of my reflections was in part influenced by the rapid time in which interviews were scheduled and held. I found it very easy to keep track of themes from interview to interview because of the short time frame between each interview. But I maintained a perspective for myself that keeping such thoughts recorded was just as important for data collection and analysis procedures at that time as it would be for later analysis when I would not be in the field and thus removed from the moment.

To prevent a loss of my evolving views of the study as much as possible, I maintained a summary table of all teacher interviews and my reflections throughout the data collection process. This table was updated following the transcription of all teacher interviews. Further, along with the use of the evolving summary table of teacher interviews, all participant involvement in either interviews or focus groups were documented into an excel spreadsheet by recording the date of contact, the date of formal participation, their name, school, and their research participant code.

All interviews and focus groups were completed by the third week of May in 2007. Days after, schools were closed for the summer. I spent the months between June and September of 2007 transcribing all data collected. The process of transcribing proved to be so time consuming that I sought again to locate someone to help with the process. However, again, I did not pursue this option due to cost limitations. I transcribed the data in order of when it was received to maintain a perspective of the context and changing focus of the study as reflected in my questions to participants. This was compared also with my summary of reflections recorded through the data collection phase. Concurrently, as I was transcribing each set of data, I kept a running list of topics or themes of interest relevant to the research questions.

At times the transcription process was so laborious that I felt removed from the data. I struggled for months to stay connected with the data. It was difficult to transcribe and allow my thoughts to progress towards an understanding of what the data were telling me. To alleviate some of this problem, I again relied on the use of the summary table of findings by using it as a checklist of what I was finding as I transcribed each set of audio

recordings. A constant cross comparison and check against my field notes and reflective entries were used to make sure that an objective and thorough use of the data was maintained.

Once all of the data were transcribed, I then found myself overwhelmed with approximately 200 pages of information. At this point, I was reminded by comments I had read from the qualitative methods research literature which emphasized the challenge of reducing the data to those elements that were most relevant for answering the research questions. At first all of the information seemed important and relevant. I relied upon the research question to assist me in keeping clear what my priorities were and were not. I also found myself having to recall the purpose of the study and the intentions I held in the beginning for how I thought the data could be used (e.g., professional development ideas) to assist in reducing the data.

While maintaining a perspective of the purpose and parameters of the study, I further engaged in the process of developing a code system for data analysis purposes. By having a combined use of an ongoing collection of topics identified through the data collection phase, and having identified specific topics of interest prior to conducting the study, I was able to create a long list of topics to pursue in the development of a coding system. This experience was detailed in the methods section as it related to inter-observer reliability checks.

While following the proposed set of procedures developed at the start of the study, I found problems in obtaining the necessary reliability in the earlier phases of the coding process. Specifically, my research assistant and I were unable to agree,

independently of each other, on what constituted a particular data segment in the transcribed data. Further, we found much disagreement on what a data segment should be coded as. This experience led me to examine and reflect on (1) the procedures for developing a coding system, (2) the lack of background experience in education of my research assistant, (3) my own limited experience in coding qualitative data, and (4) the potential for personal bias influencing the problem at hand. In all, during the course of many conversations with my research assistant we determined that perhaps all the above were likely involved. Given this, I approached my doctoral committee with this information and solicited approval to amend my procedures for data analysis as it related to coding the data, in order to solve the lack of reliability problems being found.

As with similar events discussed already, I returned to the research literature and research experts to discuss alternative approaches to working through the data. From the collective information obtained through those sources, I arrived at the conclusion that both my research assistant and myself needed to work together to (1) help support his limited knowledge of educational systems, (2) profit from his experience working with qualitative methods, and (3) minimize the impact of my personal biases or expectations. More specifically, what I found was that although I took great care in trying to remain objective and open to the responses being given by participants and even identifying unanticipated themes from interview to interview, my biases were at play when I tried to code those data.

As can be seen in the documented version of the code systems I developed (Appendix G1 to G2), I had been trying to assign unanticipated topics/themes into

categories that were developed a priori as reflected in the questions used in the teacher interview guide. Given this experience with the relative lack of experience by my research assistant in the field of education and my own limited experience coding qualitative data, it was not surprising that we were unable to find sufficient reliability in our agreement checks. To that I would say that one pattern I found that differentiated my research assistant and I was that I found he was more likely to define larger chunks of text as a data segment where I was more likely to identify smaller data segments for use.

By working together to identify data segments for all 14 interviews and both focus groups, we immediately were able to reduce a large number of disagreements. From there, we focused on condensing the number of codes being used by prioritizing what was important for answering the research questions. Thus, as can be seen between the third and fourth versions of the code system, some higher-order categories were collapsed into one because of their relatively low priority compared to other data topics. In the end, we then set about independently coding samples of the data and found we were able to reach acceptable levels of agreement.

On the issue of a manual cut and sort procedure for organizing the data, this too proved to be a challenging experience. It was time intensive and costly, and it slowed my momentum in analyzing the data. It was not until I transferred the data segments into an excel spreadsheet, along with their codes, and identifiers, that I was able to sort and organize the data in manageable ways more efficiently. The benefit of using Excel was that I was able to sort the data as I needed, either by teacher, by grade level, by topic, etc. Having organized and spent hundreds of hours with the raw transcript data, four broad

areas or themes emerged with which to summarize and report the results: (1) climate/culture in school affecting perceptions of DIBELS, (2) Supports/Resources available for teachers' use of DIBELS, (3) Teachers knowledge of DIBELS, and (4) Teachers' use of DIBELS data. This process resulted in a more evolved organization of the data (see appendix B3) and led to the development of preliminary results to be shared with stakeholders and participants.

In preparing a document for sharing the results of the study, I developed an outline of findings (Hodges, personal communication, March 2008). I confirmed the appropriateness of this while also highlighting my concerns about the length of the document. Specifically, I was concerned that the length of the document might detract some stakeholders from reviewing the results and offering feedback about the accuracy of findings. I received advice from members of my committee that by highlighting sections that specifically apply to particular groups of participants, and providing a table of contents, I could direct participants to the section that was a priority for their review. Having done this, I found 6 of the 14 teachers immediately responded back with thanks for being given the preliminary results for use. In addition, all acknowledged the accuracy and appropriateness of statements reflecting teacher perceptions.

To date, none of the specialists or reading coaches responded with any feedback. However, this was less of a concern since both groups' perceptions and comments were checked for accuracy and completeness immediately following each focus group session through a debriefing of results obtained by myself and my research assistant.

Specifically, a large note pad was used to document comments received through each

session and then used to share with the participants to confirm the information written was complete and accurate.

Attempts also were made to share results with district leaders and state leaders, specifically those involved with *Reading First*. To date, no feedback had been given nor any offered to discuss the findings. Efforts to overcome this continued at the time of this report. Attempts to understand the lack of interest by district and state leaders has led to reflections on alternative means for disseminating the findings while also considering the appropriate presentation for these audiences. Perhaps the creation of a web-site with extended links based on the organization of the findings would provide these audiences with a more user-friendly approach to view the results.

## CHAPTER FIVE – DISCUSSION

The *Reading First* grant is a core program provided through the No Child Left Behind Act (NCLB). The *Reading First* grant was initiated to promote the use of evidence-based research in reading and to develop high quality instruction in reading for kindergarten through third grade. The *Reading First* grant provided funding for three primary areas: (1) professional development for teachers, (2) purchase and implementation of reading assessments, and (3) purchase research-based curriculum and instructional materials for the classrooms. More specifically, the aim of the *Reading First* grant was to (a) increase quality and consistency of instruction in K-3 classrooms; (b) conduct timely and valid assessments of student reading growth in order to identify students experiencing reading difficulties; and (c) provide high quality, intensive interventions to help struggling readers catch up with their peers (Torgesen, 2002). At the heart of this program in Florida was the adoption of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as a required assessment for all schools receiving the *Reading First* grant to help fulfill the federal grant requirements of using assessments for screening, progress monitoring, diagnostic assessment, and outcomes evaluation.

The DIBELS (Good & Kaminski, 2002) are a set of individually administered research-based, reliable and valid measures of critical early literacy skills (Kaminski, et al., 2008). DIBELS are comprised of seven subtests for the measurement of critical early pre-reading and reading fluency skills. In Florida, five of these subtests are used. The

assessment was intended to be administered at least three times a year as a screening tool and as a means for frequently monitoring the progress of students who were assigned to receive more intensive reading instruction. Progress monitoring is a highly critical component of the intervention process with respect to determining the effectiveness of interventions for students who are at-risk for reading failure (Coyne & Harn, 2006).

Understanding how educators were using the DIBELS was considered to be a valid topic for investigation in light of the research literature on educators' use of data and the direct effects it can have on student outcomes (Fuchs & Fuchs, 2002; Stecker, Fuchs, & Fuchs, 2005; Wesson, King, & Deno, 1984; Wesson, Skiba, Sevcik, King, & Deno, 1984). At the time of this study, no studies could be found which specifically investigated teachers' perceptions and use of the DIBELS within a *Reading First* context. The present study primarily sought to understand teacher perceptions and use of the DIBELS in order to provide a foundation for exploring teacher use of DIBELS data within a *Reading First* context. Secondly, the present study sought to provide information useful for either improving current professional development initiatives for teachers, and/or identify salient variables for future research on teachers' use of DIBELS data. Three primary sources of research literature were used to provide a context regarding the discussion of the current research results: (1) data-based decision making in educational contexts (e.g. Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006); (2) Curriculum-based Measurement or CBM (e.g., Stecker, et al., 2005); and (3) system change literature (e.g., Curtis, Castillo, & Cohen, 2008).

Due to the increased accountability pressures imposed upon school districts in the United States as a function of the No Child Left Behind policies, schools have been engaging in more frequent and more complex data-based inquiry (Brunner, Fasca, Heinze, et al., 2005). Prior to the initiation of this increased climate of accountability and pressure to increase student academic outcomes, research had been ongoing since the 1970's on the utility and effective implementation of Curriculum-based Measurement (Stecker et al., 2005). Similar historical developments have occurred also in the literature concerning systems change (Curtis, et al., 2008). All combined, these sources of literature have identified several variables which may impact educators' use of data and are relevant for discussion the present study's findings. Those variables were organized similarly as the same four specific higher-order themes which emerged in the present study to provide the reader consistency in the paper's organizational format: (1) the organizational climate/culture, (2) resources available for teachers' use of data, (3) teacher knowledge and skills in using data, and (4) teacher perceptions of self and data.

#### *Discussion of Primary Results*

*Context/Culture/Climate.* Context and culture are critical for understanding when attempting to implement large-scale programs (Boyd, 1992; Curtis, et al., 2008; Fixen, et al., 2005). To begin, a body of research has indicated that administrative leadership and stakeholder involvement are essential for not only effecting the culture and climate of the school regarding the systematic use of data, but also in ensuring high quality implementation of a data-based decision making model (Allinder, 1996; Armstrong & Anthes, 2001; Bernhardt, 2000; Brunner et al., 2005; Boyd, 1992; Coburn & Talbert,

2006; Curtis, et al., 2008; Feldman & Tung, 2001; Fixen, et al., 2005; Hargreaves, 1997; Klein & Herman, 2003; Ingram, Louis, & Schroeder, 2004; Kerr et al., 2006; Lachat & Smith, 2005; Stecker, et al., 2005; Supovitz & Klein, 2003; Yell, et al., 1992; Young, 2006). In fact, Hargreaves (1997) argued most school reforms fail because efforts to exact change involve only restructuring rather than careful consideration of helping stakeholders revise their attitudes, beliefs, or perceptions about an innovation being adopted; something for which administrator involvement is essential. At the very least, stakeholders – teachers specifically – require an understanding and rationale for the change (Curtis, et al., 2008).

In the present study, attempts were made to directly investigate administrator perceptions and roles in the use of DIBELS at *Reading First* schools. However, attempts to recruit their participation were unsuccessful. It was believed that the approach used to enter the schools was insufficient to ensure administrator participation. Nonetheless, teachers were asked during interviews to comment on their perceptions of the current climate and leadership at their school regarding the use of DIBELS. On the specific topic of leadership, all teachers indicated – some more directly than others – that leadership was important for their current and continued use of DIBELS. However, there seemed to be some variability in teacher's perceptions of the amount of direct involvement a school's administrator had. Some indicated more generally that their principals support the use of DIBELS, while others indicated the principal and his/her leadership team directly review the DIBELS data often.

According to researchers leadership is important for increasing buy-in from key stakeholders when working against cultural barriers that could be significant detractors to effective data-based decision making and exacting systems change (Curtis, et al., 2008; Feldman & Tung, 2001). Though not observed directly through principal participation in the present study, it is plausible given this line of research that any variability among administrator involvement and support in the use of DIBELS may influence the variability observed in teacher perceptions of the value and utility of the DIBELS. Wayman and Stringfield (2006) summarized the importance of leadership best when they indicated it was the leaderships' role to make sure that teachers "use the data rather than be used by data" (p. 569).

Pressure to perform was a recurring theme in the present study. Several teachers voiced concerns over the increased accountability, or assessment and instruction responsibilities they were experiencing. A few of these teachers attributed these feelings of pressure directly to No Child Left Behind. This perception has been documented in previous literature (Brunner, et al., 2005; Kerr et al., 2006) and deserves some attention with respect to the present results. This climate of pressure and accountability seems to have existed in the perceptions of those who participated in the present study prior to the implementation of the DIBELS. According to some researchers (Coburn & Talbert, 2006), understanding the ways in which individual beliefs and practices are situated in and shaped by organizational context – in this case preexisting pressures of accountability and the districts reactions to the policy mandates responsible for the pressure – is important for fully understanding how teachers perceive new information or technology.

A consistent theme among teachers was the expectation that leadership was responsible for creating a positive and collaborative environment. In particular, one kindergarten teacher in the present study suggested that it was the leadership's responsibility to help deflect the pressures of the district and state policies or practices to help create a climate of collaboration and a culture of data-based decision making. This single teacher view was consistent with literature on systems change which suggests culture and context which influence increased pressures of accountability can originate not only at the individual, grade, or building level, but can also originate from the district, state, or federal level. This perspective seemed consistent with systems change literature that described the limitations and probability for failure by using a top-down approach to implementing systems change (Curtis, et al., 2008; Fullan, 1997).

According to Fullan (1997), neither the top-down or bottom-up approaches to exacting a process for systems change have found much success. Specifically, he indicated that a top-down approach often fails because schools are too complex to be controlled merely from the top (e.g., restructuring policies and procedures to mandate staff practices). This perspective is consistent with Curtis, et al., (2008) view of schools as "living organizations." They argue that schools are social systems and living organizations because they reflect the attitudes, beliefs, and practices of the people within them. As such, they argue no two schools are likely to be functionally the same given the unique characteristics of the people within each respective school. Therefore, efforts to affect district policies and procedures must be focused on the level of the school building when implementing large scale systems change efforts. It would seem, given this view,

that some flexibility is required at the district and state levels to foster a district-wide climate of support for schools as each tries to build upon their unique strengths and find growth in their areas of concern to meet the needs of their student body. A purely top-down approach is not likely to foster such flexible support for schools to find a process of implementation that works for them, specifically.

Bottom-up approaches are not likely to be successful either according to Fullen (1997). He argued such approaches often lack the necessary coordination and vision; and therefore can lead to chaos and disillusionment. Planning for change is considered a critical first step for engaging in effective systems change (Curtis & Stollar, 1996; Grimes & Tilly, 1996; Hall & Hord, 2001; 2006). Thus, a bottom-up approach will be limited in its capacity to bring about needed change through a process of finding a shared vision and involvement of key stakeholders.

Central to both of the approaches mentioned above is the lack of recognition of the importance of addressing cultural barriers, identifying cultural or contextual resources, and/or developing opportunities for involvement by all stakeholders. Returning to the results of the present study, it is not surprising given the research on systems change that teachers expressed variability in their views concerning their perceptions and understanding of using DIBELS. First, teachers were not involved directly at all levels with the adoption of the *Reading First* grant within their school. Second, a top-down approach to implementing the use of DIBELS likely influenced teacher's initial or sustained resistance for its use. Third, it appeared in the present study that teachers had few opportunities to engage in reflection about their current beliefs and

practices concerning reading development and instruction as it involved the use of DIBELS. Even those teachers who indicated a positive perception in their current use and understanding of DIBELS indicated several barriers to being able to reflect on their data and make use of the information for instructional decision making. Thus, it appeared that additional infrastructures were needed to more effectively engage in data analysis and utilization to support teacher's effective use of DIBELS.

Beyond the top-down and bottom-up approaches to systems change, Hall and Hord, (2001; 2006) argued for a horizontal perspective in which all participants in a system (i.e., district level, school building level, grade level, etc.) have the same vision and work together to implement that vision. The variability that existed among teachers' knowledge of the value and utility of DIBELS, was understandable given the limited opportunities for staff to engage in collaborative work and find a clear and shared vision. Where some hints of this occurred was in the context of teachers describing how their initial reactions to the use of DIBELS progressed from a culture of competition among colleagues to one of collaborative professional learning communities.

A promising model exemplifying an implementation approach consistent with Hord's (2006) horizontal perspective is the Florida Problem-Solving/Response to Intervention Model which was initiated at the time of the present report. The Florida RtI initiative was a program evaluation aimed at identifying the critical variables that lead to effective implementation of a problem-solving/response to intervention model of service delivery (Batsche, et al., in press). At the heart of this model is a three stage implementation model that focuses initially on developing building-wide consensus

through activities designed to increase educator knowledge and opportunity to participate in discussion about RtI and problem-solving. Following the attainment of consensus among staff and appropriate levels of buy-in among all stakeholders in the school building, implementation efforts prioritized an analysis of the school's infrastructure for supporting systematic data collection, data management, and data utilization among staff. In this stage, schools identify gaps in their current services to students and utilize data to prioritize service delivery needs as well as provide on ongoing analysis of quality improvement at the building level. The final stage of implementation articulated in the Florida RtI initiative is implementation. Overall, the Florida RtI initiative seeks to evaluate the conditions necessary for effective implementation of a problem-solving/response to intervention model by utilizing best practices in systems change efforts as described in the research literature (e.g., Hall & Hord, 2006).

Collaborative use of data among grade level teams of teachers was reported to emerge over time during the implementation of the *Reading First* grant. However, at first, many indicated a climate of competition among colleagues and a fear of retribution; both of which have been identified in the research literature as barriers to effective use of data to make educational decisions (Armstrong & Anthes, 2001; Bernhardt, 2000). Moreover, there are many identified benefits of engaging teachers in a professional learning community (Hargreaves, 1997; Hord, 1997). These benefits include a way to create a culture of change in schools, create a shared vision and purpose, build upon and learn from the collective creativity of other teachers, a supportive environment for

solving problems, and a culture of valuing the use of data to make decisions (Hord, 1997).

A first grade teacher at one school reported experiencing great difficulty and fear regarding the use of DIBELS at her school. She described the climate and culture of her school as one that was highly “political.” This description was in contrast to a different first grade teacher at another school who reported a collaborative working environment at her school. Notably, this second teacher was more representative of the other first grade teachers in the study. Further research through either a process or outcomes evaluation of the *Reading First* grant may wish to take into account the level of involvement by district and school administrators and staff perceptions of a culture of collaborative data-based decision making. Further, future research may wish to more formally evaluate how involvement or lack of involvement by district administrators has contributed to teacher perceptions of practice.

Another important systems influence which may have contributed to the variability found among teachers’ perceptions and use of the DIBELS is in regards to the level of complexity of the school district. Research suggests that complex organizations such as school districts can influence educators’ understanding of new information or technology (Coburn & Talbert, 2006). That is, given the complexity of a given system or organization, it is possible that individuals working in different parts of the system may develop contrasting sets of understandings. An example of how the context of the school district’s policies and procedures were likely impacting teachers’ perceptions and use of the DIBELS may be found in participant reactions to the volumes of assessments that are

required by the school district in addition to those required by the *Reading First* grant. Specifically, the high level of assessment activity that occurred in the participating school district served as a barrier to teachers in the following manner: (1) limited time and resources to engage in data collection, analysis and utilization using the DIBELS; (2) limited or no opportunities to engage in progress monitoring using the DIBELS; and (3) influenced teachers to view the DIBELS as “just another test they have to give.”

Implementing a large-scale program requires significant resources, time, and supports for those whose practices are to change (Curtis, et al., 2008). Findings from the present study suggest the district’s assessment schedule may have contributed as a barrier to effective implementation of the *Reading First* grant by creating limited time for teachers to engage in data analysis and utilization. One of the most frequently stated barriers to teachers’ participation in the DIBELS collection, analysis, and utilization in the present study was one of time. This was reported by teachers, Reading Coaches, and specialists who participated. It may not be reasonable to expect teachers’ to change their attitudes, beliefs, and or participation levels needed for the adoption of new technology if they are not given adequate opportunities to engage in reflective thought about their current practices as they relate to student outcomes as well as an opportunity to find value in the new technology to improve their efforts to help students reach their literacy goals. Given the relatively independent nature of school sites as living organizations capable of learning and adapting while the actions of one may not necessarily directly affect another, future research may wish to explore the manner in which the vision, mission, and purpose of the *Reading First* grant were communicated from the state to the district and

subsequently from the district to school building personnel as a means of identifying barriers to successfully building a shared vision for implementing the grant.

*Resources/Supports/Infrastructure.* In the present study, several resources were identified through interviews and focus groups to support teacher's understanding and use of DIBELS. These included the availability of a full-time Reading Coach, access to a web-based data management system for reviewing and analyzing student data through the provision of graphs, consultative efforts provided by resource personnel (e.g., Title 1 staff, school psychologists, academic diagnosticians, etc.), and access to professional development opportunities in the understanding and use of DIBELS. Research has found that several resource variables are likely to influence educators' effective use of data. Teachers may reliably use the available data if there is sufficient up-front and ongoing training and consultation on the use of data (Armstrong & Anthes, 2001; Chen, Heritage, & Lee, 2005; Fuchs, Fuchs, & Hamlett, 1989; Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Kerr et al., 2006; King et al., 1983; Love, 2004; Skiba et al., 1982; Stecker, et al., 2005; Wesson et al., 1984; Yell et al., 1992); if the data are of high quality, accurate, and immediately available for use (Coburn & Talbert, 2006; Kerr et al., 2006; Lachat & Smith, 2005; Wesson et al., 1984); if teachers have available time (Allinder, 1996; Coburn & Talbert, 2006; Lachat & Smith, 2005; Yell et al., 1992; Supovitz & Klein, 2005); and if teachers have available technology and skills for using the data utilization technology (Chen et al., 2005; Stecker, et al., 2005).

Regarding the availability and quality of training and consultation available for teachers on the use of DIBELS, it was understood as part of the larger *Reading First*

implementation plan in the school district, that all *Reading First* schools were given several full-days of training on the DIBELS and the various other components to the *Reading First* grant. Of those teachers who could recall attending this initial training, all reported unfavorable perceptions about the training and generally reported it to be unhelpful. Some were distracted by the format in which the training was provided (i.e., reading from a manual), and others perceived the information to be of little value above their current knowledge based on instructing students in reading. The Florida Center for Reading Research provided ongoing training on a variety of assessment and instructional topics aimed to support teachers in their evolving professional development as it relates to *Reading First*. However, when asked if they have attended any training outside of their school campus either in the district or through participation in statewide conferences or professional seminars, no teacher indicated any such participation.

All however, indicated receiving various types and amounts of consultation from their on-site Reading Coach. All teachers reported high value for the Reading Coach and cited multiple activities the Reading Coach was able to provide them in support of their use of DIBELS. Overall, substantial variability existed in the reporting of how teachers perceived their ability to independently interpret and use the DIBELS data. Four specific topics of interest that relate to the available research on implementing formative assessments like the DIBELS (e.g., CBM) are (1) the level of training provided to teachers on how to use and interpret formative assessments, (2) the type of consultation provided to teachers, (3) teacher formative measurement skills with or without data

analysis or evaluation skills, and (4) the infrastructure available to support teacher participation in the collection, analysis, and utilization of the data.

First, a study conducted by Wesson et al. (1984) investigated the comparative academic outcomes among students with teachers who received one of three types of training (1) an initial set of three half-day workshops at the start of the year including the provision of a manual for using CBM, followed by semi-frequent phone calls or visits by an observer; (2) a train the trainers format in which teachers from different schools were trained during a one-day training workshop and provided with a manual and instructions for teaching others in their school; or (3) a full-day initial workshop, plus monthly half-day workshops through the year, plus specific and focused training content on specific skills for training combined with opportunities to practice. Results found that the students from the third group of teacher participants made the most gains in academic performance. Students whose teachers accurately and consistently applied CBM in their classrooms made better progress than those of students in other classrooms. Wesson et al. concluded that teachers required more training than just how to measure student performance with CBM, but also direct and explicit training on how to evaluate and make use of the assessment information. Similar results have been found in other studies (King et al., 1983; Skiba et al., 1982; Wesson et al., 1984).

Additionally, work in the field of teacher training by Joyce and Showers (2002) confirms the results found in preparing teachers to adopt new skills. Specifically, Joyce and Showers indicated four components to effective teacher training: (1) knowledge and understanding of the theory and rationale for the new innovation; (2) modeling of the new

skill(s) with multiple exemplars; (3) practice of the new skills in a safe environment with repeated opportunities to rehearse and receive feedback; and (4) ongoing supportive coaching or opportunities to collaborate with peers regarding implementation of the new skill(s). Moreover, the work by Little (1997) suggests the above model is insufficient to sustain teacher use of a new skill, let alone develop a culture of change. He recommended identifying the level of understanding where teachers were operating in relation to a given topic or skill being learned and tailoring the instruction or activities so the new ideas can be fully integrated into the current practices. Further, he advocated that teachers have an opportunity to express dissent and challenge prevailing beliefs and practices. He suggested that teachers do require an account of the “big picture” on educating children and the purpose of education; similar to engaging them in a “systems thinking” according to Senge’s (1990) five factors that individuals collectively need to become and function as a learning organization.

A full-time Reading Coach was available to providing coaching and ongoing professional development to teachers on the use of DIBELS. However, teachers reported high degrees of dissatisfaction with the manner in which early training was provided. Given the limited opportunities for teachers to participate as stakeholder partners in the change process, and the barriers to advancing their knowledge and understanding of DIBELS facilitated by low value perception for trainings on the topic, it is not surprising that variability was found in teacher’s ability to not only use the DIBELS data more effectively, but also value its use for instructional planning. And though the Reading Coach was consistently identified as a crucial variable towards the positive perceptions of

DIBELS among some teachers, little evidence was found which demonstrated efforts by the Reading Coaches or the school district or school buildings to build a capacity among school staff for sustainability of using the DIBELS. According to Adelman & Taylor (2003) large scale programs often fail in their implementation when temporary funding has ended and the staff hired to support implementation efforts are removed. This is a topic most central to a failure of creating the infrastructure needed to support the continued use of the program and all its components (Adelman & Taylor, 2003; Curtis, et al., 2008).

Next, regarding the type of consultation provided to teachers, researchers have found that teachers require ongoing consultative support in the development of skills and in support of using CBM data to inform instructional decision making (Fuchs et al., 1992). This research even found that teachers who reported positive and high value in the use of CBM procedures in their class still had great difficulty regarding how to interpret and make sense of the data for instructional decision making. Research like that of Fuchs et al., in response to such observations, have developed evaluation guidelines to help teachers know when to consider adjusting instruction and how to determine what to change. However, a vital pre-requisite to using such guidelines is the active implementation of progress monitoring data collection procedures; something that the State of Florida *Reading First* model only just recently started to introduce to schools rather than at the beginning of implementation.

In this context, the present study found moderate to substantial variability in the reporting among teachers regarding the frequency and types of support received by their

Reading Coach. It cannot be determined from the present study the degree to which Reading Coaches provided teachers explicit consultation and training guidance towards the independent use and analysis of their own data. However, results consistently found perceptions among teacher and non-teacher participants that teachers were still highly dependent on and in need of continued supports to further their development of skills in the analysis and interpretation of DIBELS data. Future process or outcome evaluations of the *Reading First* grant should consider the frequency and quality of consultative efforts provided by Reading Coaches when evaluating the impact of teachers' use of DIBELS data in relation to student academic outcomes in reading.

Regarding the topics of training, none of the teachers could recall receiving any training, with exception to what was provided by their Reading Coach, on how to independently analyze and link their class data to instructional decision-making. Some voiced strong interest in more training to learn how best to do this. The Florida Center for Reading Research provided several online resources and guides to assist teachers in this context, but none of the teachers interviewed were aware of such information; some were not even aware of the website beyond the availability of the PMRN database. To this researcher the Reading Coach was highly valuable in supporting the school staff in the collection, interpretation, and utilization of DIBELS data. However, it appeared from the teacher interviews that teachers were highly dependent on the Reading Coach for this continued set of supports and would likely experience substantial difficulties without such support when the grant expires. Indeed, several researchers have noted that of the various influences that exists on educators' use of data, data analysis, interpretation, and

utility for decision making is one of the most elusive areas for professional development initiatives (Brunner et al., 2005; Casey, Deno, Marston, & Skiba, 1988; Cizek, 2001; Fuchs et al, 1992; Fuchs, et al., 1989; Fuchs & Fuchs, 1986; 2002; Grigg, Snell, & Loyd, 1989; Herman & Gribbons, 2001; Kerr et al., 2006; Lachat & Smith, 2005; Stecker, et al., 2005; Tindal et al., 1981; Wesson, 1991; Young, 2006).

The implications of this professional development need and support for teachers is highly relevant to the use of DIBELS. For example Fuchs, et al. (1989) investigated the academic outcomes of students from teachers who participated in one of three groups: (1) CBM measurement use plus evaluation consultation; (2) CBM measurement only without evaluation support, though receiving consultation as general a support for using CBM; and (3) a control group that did not use CBM. Results found that the students from the class in which CBM measurement was used plus consultation and training on how to use the information to make decisions out-performed students from the other two groups on CBM outcome measures. Of particular interest was their observation that teachers who collected CBM data, but did not make use of that data did not effect any better student outcomes than those in control conditions where CBM measures were not used. Thus, any process or outcome evaluations of the *Reading First* grant must consider the student outcomes in context of how effectively and frequently teachers were using the data to make frequent and targeted instructional changes in their classes.

Related to the topic of utilizing data for instructional planning is the use of DIBELS for screening, progress monitoring, diagnosing academic needs, and evaluating student outcomes. Coyne and Harn (2006) provided a conceptual framework for thinking

about early literacy assessment across the four distinct assessment purposes just listed above. Recall, that in the present study all teachers described the DIBELS as a test that is administered three times a year; thus characterizing the use of DIBELS as a screening tool for identifying students who may be at-risk for reading difficulties. Although several teachers reported using the DIBELS to identify students in need of additional supports or interventions, none of the teachers reported using DIBELS for progress monitoring. However, two teachers indicated knowledge of the option of using DIBELS as a progress monitoring tool, but cited a lack of access to either necessary materials or training to use DIBELS for progress monitoring.

Of particular interest was the observation that many of the teacher who participated in the present study demonstrated data analysis skills consistent with the DIBELS experts. It is important to point out that these teachers had also voiced positive perceptions and value in using DIBELS data as well as reported relatively higher self-efficacy in using the DIBELS data to organize learning groups and assign standard interventions for student in need of assistance. Further, it is important to note that none of the teachers had been engaged in any progress monitoring activities with the DIBELS at the time of the study; thus, no such data were used in the case studies for teacher use.

According to Coyne and Harn (2006), progress monitoring allows teachers to determine if students are making adequate growth toward meeting grade-level reading outcomes. Further, they suggest as best practice that the DIBELS be frequently used for formative evaluation with students receiving interventions to monitor their progress and measure the effectiveness of interventions. This point is critical from this researcher's

perspective. First, as Coyne and Harn have indicated, screening and progress monitoring activities alone do not tell an educator everything that is needed to make instruction more effective or efficient. This was consistent with interview comments provided by the DIBELS experts in the present study. However, progress monitoring does provide a pulse, or “vital sign,” as stated by Coyne and Harn, for the teacher to track, in a reliable, and non-intrusive manner, how a student is functioning in the class on a particular set of skills and whether an intervention is having the desired impact.

Some teachers in the present study reported concern over the amount of time between DIBELS benchmark cycles as a barrier to using the data to make instructional decisions. So, without either the knowledge of using DIBELS for progress monitoring, having access to progress monitoring materials, and/or training on how to administer DIBELS, teachers are left with a void with which they filled by using either less technically adequate assessment tools for monitoring student progress, or their own personal judgments to assess students’ responses to instruction.

And yet, before one considers the implication this all has on professional development needs in the participating district, recall that many teachers, when asked if they would like to learn how to administer the DIBELS, rejected the idea citing overwhelming assessment and instructional responsibilities. It would seem appropriate that any program evaluation of the *Reading First* model take into consideration the degree to which students are not being monitored either at all, or are being monitored with assessments that have a broad assessment focus at best, or lack technically adequate data at worst.

A final area of consideration regarding variables that can influence teacher's use of data were teachers' perceptions and attitudes concerning the use of a particular assessment. Research has suggested teacher's use of data can be influenced by teachers' perceived validity of an assessment tool (Coburn & Talbert, 2006; Kerr et al., 2006; Yell et al., 1992); staff attitudes about using data to inform practice (Bernhardt, 2000; Ingram et al., 2004; Lachat & Smith, 2005); teachers' self-efficacy in supporting student learning (Allinder, 1995; 1996; Ingram et al., 2004); teachers' response to change (Allinder, 1996; Yell et al., 1992); teacher perceptions of new initiatives as additional work (Wesson et al., 1988; Yell et al., 1992); and beliefs about what constitutes data (Allinder, 1996; Coburn & Talbert, 2006; Ingram et al., 2004; Farlow & Snell, 1989; Young, 2006),

In the present study, teacher perceptions about the value and utility of the DIBELS were substantially variable across teachers and schools. Several teachers characterized the DIBELS as an invalid test because (a) the tasks are timed, (b) the tests are conducted by someone other than the teacher, (c) it is given a higher priority by their administrators over assessments the teachers have long used and with which they have become comfortable (e.g., Running Records), (d) the results did not correspond with other assessments or observations conducted by the teachers in the classroom, and/or (e) it uses measures that they perceived are "inappropriate" for their grade level (e.g., NWF subtest). The comments by teachers about DIBELS as a "timed test" suggest that many teachers see the DIBELS as a high stakes test that students are required to pass – for which teachers feared professional evaluations either early in the use of DIBELS or even presently as one teacher reported.

Reading Coaches, specialists, and DIBELS experts in the present study had reported the above concerns. It would seem plausible that one professional development option is to share with teachers the psychometric properties of the DIBELS in the context of developing the test and that doing so would serve as a foundation for helping teachers to see the valid use of DIBELS. As one teacher and several Reading Coaches indicated in the present study, they have found teacher perceptions change in favor, or at least greater acceptance for, the DIBELS when it is shared that research has found a strong correlation between end-of-first grade ORF measures and later third grade FCAT performance.

When Reading Coaches were asked during a focus group to comment on teachers' use of data, they indicated that what was essential was a culture of valuing and using data. Teachers, who were interviewed however, were observed to not necessarily value data – though some did explicitly state such a value for DIBELS over other assessments – but suggested a culture of collecting data for the sake of collecting data. Teachers overwhelmingly communicated that there is too much assessment being conducted in the schools – and at times to the detriment of protecting instructional times. They perceived the DIBELS to be just one more such assessment; much like Yell et al.'s discussion of CBM perceived as an “add-on” by teachers as a barrier to effective implementation and use.

In this same context, almost all teachers spoke of having insufficient time to plan, research alternative instructional ideas, consult with other teachers or the Reading Coach, or to even review, let alone analyze, their class DIBELS data. Having been an employee

in the present district, it has been observed first hand that the amount of testing that teachers are required to participate in is extensive. It seems unlikely that professional development activities provided to teachers on what the DIBELS are, how to administer the measures and use the data will be enough to overcome these issues. Rather, such training would likely be limited in value and effectiveness to the degree to which teachers must continue to juggle their time around the multiple assessment schedules that impact their instructional times and/or take away from their time to plan and use data to guide their efforts.

### *Limitations*

Having discussed the results of the present report in light of the relevant and available research, discussion of several limitations to the study is warranted. First, the study was conducted primarily to provide a rich description of teacher perceptions and use of the DIBELS to understand the conceptions among individuals participating in a large-scale grant program within a school district – something that Coburn & Talbert (2006) have indicated little research has undertaken. As such, only a small sample of perceptions was obtained that may exist on the chosen topic. Given the limitations, the results do not lend themselves to generalization across the school district necessarily, and certainly not across the state. Future research may wish to conduct similar research studies either within other school districts receiving the *Reading First* grant, or across several school districts receiving *Reading First* grants to evaluate the consistency of findings among other sites and participants.

Another limitation in the present study may be found in the recruitment procedures used in the present study. Though a saturation method was used to determine the extent to which possible unique perspectives were found, the way in which teachers were selected may limit the degree to which the findings are representative of the teachers in the participating school district. Due to constraints imposed by the school district's Institutional Review Board, principals had to be informed of teachers being recruited for participation and principals had to give their permission for teachers at their school to participate. These constraints may have contributed to the low interest found among teachers to participate in the study. Also, teachers essentially self-selected themselves to participate in the study and therefore may not be completely representative of all teachers in the district.

One possible safeguard against this limitation was in the use of a saturation method (Patton, 2002) to identify the point at which data collected no longer yielded any new information. Application of the saturation method led the recruitment of only 14 teachers having found no new information from the final two interviews for each grade level, respectively. Nonetheless, future research should consider larger-scale applications of the present study's approach to answering the research questions to identify any salient perceptions not captured in the present study.

Similar limitations also existed in the recruitment of Reading Coaches and Specialists. Little variability was found within each group of responses. All who participated in either the Reading Coaches or specialists focus groups voiced unanimous value in the use of DIBELS. It cannot be determined if the high degree of consistency

among these two separate groups was a function of either (1) a higher knowledge and understanding of the DIBELS when compared to teachers, (2) a non-representation of other Reading Coaches and specialists given the self-selection approach to recruitment, or (3) the focus group format may have resulted in contrasting perceptions being withheld by participants out of concern for sharing opposing views in the group format. To safeguard against the third possibility, all participants in each group were given the primary researcher's contact information and encouraged to contact him if they wanted to share any additional information with the understanding that all their information would be kept confidential. No participants from either focus group engaged the researcher in additional conversation on the study's topic outside of the focus groups.

Another limitation to the findings was related to the fact that for the teachers who were interviewed, the Reading Coach or specialists at their school may not have participated. Thus, any input provided by Reading Coaches and specialists cannot be extracted to investigate specific claims or observations made by a specific teacher at the same school. However, because data from the teachers were aggregated within each group, it was possible to explore the general themes that arose through teacher interviews and provide a comparative analysis against the views of Reading Coaches, DIBELS Experts, and specialists. Additionally, the present study sought to document teacher perceptions and use of the DIBELS, and not an investigation of their accuracy of those perceptions against actual events occurring in their school by others.

Another limitation was the limited resources available to the researcher to provide a quick provision of the results to identified stakeholders. They delay in providing the

results may have negatively impacted their motivation to provide any feedback or discussion of the findings. At the time of the present report, efforts continued to engage relevant stakeholders at the state and local level to discuss the results and their possible implications for practice.

Another limitation involves the researcher's limited experience in conducting and participating in qualitative research. However, several activities were engaged in to minimize this potential weakness. First, the involvement of a doctoral committee provided guidance and review of the research process to ensure that an appropriate study was being conducted that would be valuable to the research literature, that it involved sufficient methodological rigor, and that it was conducted in a reliable manner. Second, additional researchers with known expertise in qualitative methodology were consulted regarding the methodologies and practices of the present study to ensure that the researcher was conducting an appropriate and reliable study. Third, several quality assurance methods were employed to provide a credible account of the research topic investigated. These methods included (1) a rich and thick description of the research process, (2) documentation of the researcher's position and biography, (3) saturation of data, (4) purposeful search for variation and maximization of responses, (5) use of a triangulation across data sources, (6) inter-observer reliability checks, (7) member checks for accuracy and completeness of data, and (8) peer review by members of the doctoral committee.

A final limitation may be found in the selection of kindergarten and first grade teachers only. The decision to limit the study to these grade levels was because the

DIBELS subtests used for benchmark assessments at those grade levels comprise the entire DIBELS assessment used in Florida. Higher grade levels were not investigated in the present study. Thus, the findings of this study should not be generalized with respect to implications for supporting teachers in those higher grade levels as it relates to DIBELS.

#### *Contributions of Present Study*

Despite the identified limitations of the present study, several strengths or contributions are worth discussing. At the start of conducting the study, little research could be found which investigated teacher perceptions and skills regarding the use of DIBELS within a *Reading First* context. This study provided a foundation for future qualitative and mixed-methodological approaches to investigating teacher use of DIBELS and evaluation of the process of implementing and maintaining the *Reading First* model of reading as it related to the use of DIBELS, or similar assessments.

Next, the present study identified variables and topics worthy of continued investigation as found in related fields of study such as systems change, data utilization, and best practices in the use of DIBELS. Earlier research on teachers' use of CBM-Reading assessments to improve student outcomes, coupled with best practice knowledge of effective systems change and the related variables that can impact schools' use of data converged in the present study to provide a broad context for improving efforts towards: (1) professional development for teachers in the effective use of DIBELS, (2) building capacity and infrastructure for schools to support teachers' use of DIBELS, (3) identifying opportunities to allow teachers to participate as stakeholder participants in the

continued school efforts and to support implemented features of the *Reading First* grant, and (4) building a rationale for the school district and building administrators to reanalyze their shared vision and efforts to create and foster a culture that values data utilization as well as collaborative planning and problem-solving.

### *Final Summary*

The present study was conducted to understand teacher perceptions and use of the DIBELS as a core component of the overall *Reading First* model in the State of Florida. Several barriers and resources were discovered which hindered or fostered effective adoption and use of DIBELS by educators, respectively. Given the results of the present study, and having provided a context for interpreting the results within a larger body of research in systems change, data utilization, and best practices in using CBM/DIBELS, several recommendations are appropriate for those interested in implementing the use of DIBELS at their school or school district:

#### *1. Create a culture of change and shared vision.*

- a. Provide educators with a conceptual model for using data and solving problems at the building, grade, classroom, and student levels. The authors of the DIBELS recommend the use of DIBELS within the context of the Outcomes-Driven Model. Additionally, the recent legislative requests for schools to adopt a Response to Intervention (RtI) approach to service delivery provides such a framework for helping schools learn and adopt a problem-solving model for effective data management and utilization. At the time of this report the Florida RtI initiative was being implemented and is consistent with this current

recommendation. Evidence from that initiative will be important for review in light of this recommendation.

- b. Ensure leadership at the school building level is directly and consistently involved in implementing the use of DIBELS and utilizing data to inform instruction.
- c. Encourage a shared vision for the use of formative assessments among staff and a culture that embraces data analysis and data-based decision making.
- d. The school district may wish to revise its current district assessment plan to consider using formative assessments similar to DIBELS that have documented reliability and validity while also affording time and cost efficient means of collecting student data with minimal interruption of class instruction.
- e. The school district may wish to consider replacing their district reading assessments for all elementary schools with the DIBELS, CBM reading, or similar formative reading assessments with sufficient technical adequacy for the purpose of screening and progress monitoring; a summative district assessment could still be provided at the end of the year. This change would afford the opportunity to use DIBELS, or measures like them, more effectively for progress monitoring. Additional testing would be restricted to those cases where more diagnostic information is needed for intervention design.
- f. Teachers are the most important stakeholder when implementing the use of DIBELS or any other formative assessment system. It will be key to involve them in all aspects of decision making and implementation design.

- g. Schools may wish to consider the establishment of a school-based team comprised of one administrator, representatives for all grade levels, and representatives from ESE and Student Services as a venue for regularly analyzing school data, supporting teacher professional development needs, and developing plans to increase student outcomes in reading.
  - h. Principals and district administrators, as well as support personnel like school psychologists, should be familiar with literature on effective systems change while also becoming familiar with recognizing teachers' level of concern (Hall & Horn, 2006) in response to change so as to tailor any trainings or professional development activities to effectively support teachers in the analysis and use of data.
  - i. The school district may wish to consider either discontinuing the use of assessments that do not provide opportunities for instructional decision making, or provide supports to schools to have them completed in a timely manner with minimal impact on instructional times.
2. *Create Capacity and Infrastructure to Support Use of DIBELS*
- a. Teachers need more time to engage in data analysis and utilization. This can be achieved by prioritizing the types and frequencies of assessments that are required for collection through a RtI/Problem-Solving model that increases assessment density only as a function of severity of student underachievement and/or need for more information to identify successful interventions.

- b. The school district and school building administrators should consider making the role of the Reading Coach permanent or identify a staff person who can take on the full-time responsibility of facilitating systems change and providing ongoing professional development to teachers on the effective use of DIBELS data.
- c. Teachers need access to color printers if they are expected to effectively use the PMRN graphs.
- d. The professional learning community format for grade level teacher teams being used by the school district should be maintained and supported as an appropriate venue for collaborative problem-solving among teachers.
- e. Teachers need ongoing support and consultation to maintain and establish greater skills in data analysis and utilization.
- f. DIBELS assessors should be encouraged to provide notes and/or record specific student errors so that teachers can more effectively evaluate error patterns and make use of the information for intervention development.
- g. Continue to encourage teachers to use the scoring protocols for analysis of errors towards the development of either intervention plans or consideration of other diagnostic measures.
- h. Infrastructures need to be developed to support ongoing progress monitoring of students identified as being at-risk for later reading difficulties. This recommendation should be accomplished within the context of adopting a conceptual problem-solving model (i.e., the Outcomes-Driven Model or RTI) for

allocating resources, prioritizing goals, and providing professional development opportunities for teachers.

*3. Provide Additional and Ongoing Training on the Use of DIBELS/Formative*

*Assessments*

- a. Provide teachers with a broad “big picture” rationale for why DIBELS is an effective method for screening, monitoring progress, and evaluating effective reading programs. This may be accomplished by providing the historical development and case study examples that exist in the literature on the use of DIBELS and other similar formative assessments.
- b. Some teachers may require additional knowledge on what DIBELS measures and how it corresponds with the development of reading skills. They also require knowledge on how DIBELS differs as a General Outcome Measure as compared to the mastery-oriented assessments used through the school district.
- c. Teachers need to further understand what formative assessments offer for differentiating instruction.
- d. All educators need to understand the difference between benchmark assessment and progress monitoring and the value and purpose of both.
- e. Educators need to understand the importance of progress monitoring and staff need to be held accountable and supported by building principals and district leaders to engage in progress monitoring activities using formative assessments like the DIBELS.

- f. Teacher would benefit from learning about research on the development of DIBELS and the establishment of benchmark goals (e.g., end of year expectations). This information will help them understand conceptually what the benchmark standards are and how they were developed so they see the correlation those benchmark standards have with successful reading in upper grade levels.
  - g. Teachers need support and further training on the importance of fluency of early literacy skills and their links to comprehension. The concept of fluency and general outcome indicators are critical for teachers if they are to avoid perceiving the DIBELS as a “speed test.”
  - h. Some teachers may need additional training and support to help them understand the importance of using nonsense words as a measure of student decoding skills and the value of measuring such skills.
4. *Provide Educators with Ongoing Supports for Data Analysis and Utilization*
- a. Ensure that data is collected in a timely manner and immediately available for teachers (within a week).
  - b. Educators at all levels of the school system require ongoing support and training for analyzing graphs/data, particularly the PMRN if that system is to be implemented along with the use of DIBELS. Schools have become effective and efficient in collecting data; however, they are often missing a conceptual and knowledgeable approach to analyzing the data. Both the Outcomes-Driven Model and the RtI/Problem-Solving model for data-based decision making provide a

conceptual approach for schools to analyze their data to inform their practices as well as evaluate their instructional and programmatic efforts.

- c. Teachers need to be encouraged, allowed time, and supported in their efforts to access and use the PMRN.
- d. Provide explicit and systematic training for teachers on how to use PMRN reports to answer questions about level of intervention focus within a tiered system of service delivery like that being offered currently within a Response to Intervention (RtI) model (Castillo, Cohen, & Curtis, 2006).
- e. Teachers should be encouraged and or trained to observe and prioritize student performance trends rather than single snapshots of data. Examining performance in this manner requires the collection of progress monitoring data to establish reliable trends for analysis.
- f. Provide ongoing support, training, and expectations that schools and teachers use multiple sources of data when making decisions about students. Further, training should focus on how to reconcile data sources that are seemingly in contrast with each other (i.e., acceptable performance on one measure and not another).
- g. Schools need to prioritize the use of ongoing progress monitoring. The lack of such monitoring appears to be driven by low consensus among staff for the value of monitoring progress more frequently coupled with insufficient infrastructure for supporting this level of data collection. In addition, the decision to not implement ongoing progress monitoring activities from the start of the *Reading*

*First* grant in Florida also may have influenced educators' views with respect to progress monitoring.

In closing, the present research study was conducted to understand the perceptions and uses of the DIBELS from the views of classroom teachers and non-teaching educators in an effort to identify salient variables for later research as well as provide targets of opportunity for later professional development initiatives to support teachers' use of DIBELS or similar formative assessments. It was hoped, with continued focus and research on the variables that increase the effectiveness and understanding of how educators implement and use formative evaluation systems, that schools will be able to find greater efficiency and effectiveness towards improving student academic outcomes in reading.

## List of References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation, 14*, 1-25.
- Allinder, R. (1995). An examination of the relationship between teacher efficacy and curriculum-based measurement and student achievement. *Remedial and Special Education, 16*, 247-254.
- Allinder, R. (1996). When some is not better than none: Effects of differential implementation of curriculum-based measurement. *Exceptional Children, 62*, 525-535.
- Armstrong, J., & Anthes, K. (2001). How data can help: Putting information to work to raise student achievement. *American School Board Journal, 188*, 38-41.
- Baker, S., Smith, S., Kame'enui, E. J., McDonnell, M., & Gallop, S. M. (1999). *A blueprint for bridging the gap between research and practice involving the Springfield Public Schools and the University of Oregon*. (Final Report, U.S. Department of Education Grant H023G50021). Washington, D.C.: U.S. Department of Education.
- 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. M., Curtis, M. J., Dorman, C., Castillo, J. M., & Porter, L. J. (In press). The Florida Problem-Solving/Response to Intervention Model: Implementing a statewide initiative. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.
- Battistich, V., Schaps, E., Watson, M., & Solomon, D. (1996). Prevention effects of the Child Development Project: Early findings from an ongoing multisite demonstration trial. *Journal of Adolescent Research, 11*, 12-35.
- Bernhardt, V. (2000). Intersections. *Journal of Staff Development, 21*, 33-36.

- Blakely, C., Mayer, J., Gottschalk, R., Schmitt, N., Davidson, W., Roitman, D., & Emshoff, J. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253-268.
- Boyd, V. (1992). *School context: Bridge or barrier for change*. Austin, TX: Southwest Educational Development Laboratory.
- Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mandinach, E., & Wexler, D. (2005). Linking data and learning: The grow network study. *Journal of Education for Students Placed At Risk, 10*, 241-267.
- Buck, J. & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test*. (FCRR Technical Report #1) Tallahassee, FL: Florida Center for Reading Research.
- Casey, A., Deno, S., Marston, D., & Skiba, R. (1988). Exceptional teaching: Changing teacher beliefs about effective instructional practices. *Teacher Education and Special Education, 11*, 123-132.
- Castillo, J. M., Cohen, R., & Curtis, M. J. (2006). Evaluating intervention outcomes: A problem-solving/Response to intervention model as systems-level change. *Communique', 35*, 8.
- Castro, F. G., Barrera, M. Jr., & Martinez, C. R. Jr., (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*, 41-45.
- Center for Substance Abuse Prevention (CSAP), (2001). *Finding the balance: Program fidelity and adaptation in substance abuse*. Rockville, MD: SAMHSA, U.S. Department of Health and Human Services.
- Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students' learning needs with technology. *Journal of Education for Students Placed At Risk, 10*, 309-332.
- Chen, H. and Rossi, P. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.
- Coburn, C., & Talbert, J. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*, 469-495.

- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton-Mifflin Company.
- Coyne, M., & Harn, B. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools, 43*, 33-43.
- Cizek, G. (2001). Conjectures on the rise and fall of standards setting: An introduction to context and practice. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Curtis, M. J., Castillo, J. M., & Cohen, R. M. (2008). Best practices in system-level change. In Thomas, A., & Grimes, J. (Eds.) *Best practices in school psychology V*. (pp. 887-902). Washington, D.C.: National Association of School Psychologists.
- Curtis, M. J., & Stollar, S. A. (1996). Applying principles and practices of organizational change to school reform. *School Psychology Review, 25*, 409-417.
- DePaulo, P. (2000). Sample size for qualitative research. *Quirk's Marketing Research Review, 636*, 1-5. Retrieved March 20, 2006, [www.quirks.com/articles](http://www.quirks.com/articles).
- Durlak, J. A. & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25*, 115-152.
- Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*, 47-53.
- Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills-Modified. *School Psychology Review, 30*, 33-49.
- Farlow, L. and Snell, M. (1989). Teacher use of student performance data to make instructional decisions: practices in programs for students with moderate to profound disabilities. *Journal of the Association for Persons with Severe Handicaps, 14*, 13-22.
- Feldman, J., & Tung, R. (2001). Using data-based inquiry and decision-making to improve instruction. *ERS Spectrum, 19*, 10-19.

- Fielding, N. G. (1993). *Qualitative data analysis with a computer: Recent developments*. Social Research Update, March, University of Surrey.
- Fitzgerald, T. P. & Clark, R. M. (1976). Process evaluation for inservice training. *Reading Improvement, 13*, 194-198.
- Fixen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication 231). Tampa, FL: Louis de la Parte Florida Mental Health Institute, University of South Florida.
- Fletcher, J. M., Francis, D. J., Morris, R. D., Lyon, G R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 506-522.
- Florida Center for Reading Research (FCRR), (n.d.). Progress Monitoring and Reporting Network. Retrieved March 20, 2006, from [www.fcrr.org](http://www.fcrr.org).
- Florida Department of Education (n.d.) *Just Read Florida!* Retrieved March 20, 2006, from [www.justreadflorida.com/docs/guidance.pdf](http://www.justreadflorida.com/docs/guidance.pdf).
- Foorman, B. R., Francis, D. J., Fletcher, J. M., Paras, M., & Schatschneider, C. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37-55.
- Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry, 36*, 717-732.
- Fuchs, L., Deno, S., & Mirkin, P. (1984). The effects of frequent curriculum-based measurement and evaluation on student achievement, pedagogy, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.
- Fuchs, L. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199-208.
- Fuchs, L & Fuchs, D. (2002). Curriculum-based Measurement: Describing competence, enhancing outcomes, evaluating treatment effects and identifying treatment nonresponders. *Peabody Journal of Education, 77*, 64-84.
- Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of instrumental use of Curriculum-based Measurement to enhance instructional programs. *Remedial and Special Education, 10*, 43-52.

- Fuchs, L., Fuchs, D., Hamlett, C., & Ferguson, C. (1992). Effects of expert system consultation within Curriculum-based measurement, using a reading maze task. *Exceptional Children, 58*, 436-450.
- Fuchs, L. S., Wesson, C., Tindal, G., Mirkin, P., & Deno, S.L. (1982). *Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures*. Research Paper #64. University of Minnesota: MINN. Institute for Research on Learning Disabilities.
- Fullan, M. (1997). *The challenge of school change*. Arlington Heights, IL: IRI/SkyLight Training.
- Glasser, B. G. & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Good, R. H., Gruba, J., & Kaminski, R. A. (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 699-720). Washington, D.C.: National Association of School Psychologists.
- Good, R. H. & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11*, 326-336.
- Good, R. H., & Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency passages for first through third grades* (Technical Report No. 10). Eugene, OR: University of Oregon.
- Good, R. H. Kaminski, R. A., Smith, S., Simmons, D. S., Kame'enui, E. J., & Wallin, J. (2003). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: System for observing teaching and learning*. Baltimore: Paul H. Bookes.
- Good, R. H., Kaminski, R. A., Simmons, D., Kame'enui, E. J. (2001). Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an Outcomes-Driven Model: Steps to Reading Outcomes. *OCCS Bulletin, 44*, Win 2001.
- Good, R. H., Simmons, D. S., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

- Good, R. H., Simmons, D. S., Kame'enui, E. J., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.
- Good, R. H., Simmons, D. S., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27*, 45-56.
- Goodstadt, M. S., (1988). School-based drug education in North America: What is wrong? What can be done? *Journal of School Health, 56*, 278-281.
- Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (p. 467-564). Mahawah, NJ: Lawrence Erlbaum Associates, Inc.
- Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science, 12*, 1-27.
- Griffiths, A. J., VanDerHeyden, A. M., Parson, L. B., Burns, M. K. (2006). Practical applications of response to intervention research. *Assessment for Effective Intervention, 32*, 50-57.
- Grigg, N., Snell, M., & Loyd, B. (1989). Visual analysis of student evaluation data: A qualitative analysis of teacher decision making. *The Journal of the Association for Persons with Severe Handicaps, 14*, 23-32.
- Grimes, J. & Tilly, W. D. (1996). Policy and process: Means to lasting educational change. *School Psychology Review, 25*, 465-476.
- Hall, G. E. & Hord, S. M. (2006). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.
- Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.
- Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., Fleming, C. B. (1999). Opening the black-box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology, 27*, 711-731.
- Hargreaves, A. (1997). *Rethinking educational change with heart and mind. 1997 ASCD Yearbook*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Helitzer, D. L., Davis, S. L., Gittelsohn, J., Gohn, S., Murray, D. M., Snyder, P., & Stecker, A. B. (1999). Process evaluation in a multi-site obesity primary prevention trial for Native American school children. *Supplement to the American Journal of Clinical Nutrition*, 69, 816S-824S.
- Helitzer, D., Yoon, S., Wallerstein, N., Dow y Garcia-Velarde, L. (2000). The role of process evaluation in the training of facilitators for an adolescent health education program. *Journal of School Health*, 70, 141-147.
- Herman, J., & Gibbons, B. (2001). Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation (CSE Tech. Rep. No. 535). Los Angeles: University of California, Center for the Study of Evaluation.
- Hintze, J. M., Ryan, A. L., and Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32, 541-556.
- Hord, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Southwest Educational Development Laboratory: Austin, TX.
- Individuals with Disabilities Education Improvement Act of 2004, Pub. L. 108-466.
- Ingram, D., Louis, K., & Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106, 1258-1287.
- Israel, B., Cummings, K., & Dignan, M. (1995). Evaluation of health education programs: current assessment and future directions. *Health Education Quarterly*, 22, 364-389.
- Johnston, P. & Allington, R. (1991). Remediation. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp.984-1012). New York: Longman.
- Joyce, B., & Showers, B. (2002). *Student achievement through staff development*, 3<sup>rd</sup> Edition. Alexandria, VA: The Association for Supervision and Curriculum Development.

- Kaminski, R. A., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1181-1204). Bethesda, MD: National Association of School Psychologists.
- Kaminski, R. A. & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215-227.
- Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M.R. Shinn (Ed.) *Advanced applications of curriculum-based measurement* (pp. 113-142). New York: Guilford Press.
- Kame'enui, E. J., & Simmons, D. C. (2000). Beyond effective practices to schools as host environments: Building and sustaining a school-wide intervention model for reading. *OSSC Bulletin*, 41, 3-24.
- Kerr, K., Marsh, J., Ikemoto, G., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112, 496-520.
- King, R. P, Deno, S., Mirkin, P., & Wesson, C. (1983). *The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison*. Research Paper #111. University of Minnesota: MINN. Institute for Research on Learning Disabilities.
- Klein, J. & Herman, R. (2003). The contribution of a decision support system to educational decision-making processes. *Journal of Educational Computing Research*, 28, 273-290.
- Kovaleski, J. F., Gickling, E. E., Morrow, H., & Swank, P. R. (1999). High versus low implementation of instructional support teams: A case for maintaining program fidelity. *Remedial and Special Education*, 20, 170-183.
- Lachat, M. & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Places At Risk*, 10, 333-349.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications, Inc.
- Little, J. W. (1997). Teachers' professional development in a climate of educational reform. In M. Fullan (Ed.), *The Challenge of School Change* (pp.57-84). Arlington Heights, IL: Skylight Professional Development.

- Love, N. (2004). Taking data to new depths. *Journal of Staff Development*, 25, 22-46.
- Marshall, C. & Rossman, G. B. (1999). *Designing qualitative research*. Third Edition. Thousand Oaks: Sage.
- Maxwell, J. (2005). *Qualitative research design: An interactive approach*. Second Edition. Thousand Oaks, CA: Sage Publications.
- McGraw, S., Sellers, D., Stone, E., Bebachuk, J., Edmundson, E., Johnson, C., Bachman, K., & Luepker, R. (1996). Using Process Evaluation to Explain Outcomes. *Evaluation Review*, 20, 291-312.
- Merriam, S. B., (2002a). Introduction to qualitative research. In Merriam and Associates (Eds.) *Qualitative research in practice: Examples for discussion and analysis* (pp. 3-17). San Francisco, CA: Jossey-Bass.
- Merriam, S. B., (2002b). Assessing and evaluating qualitative research. In Merriam and Associates (Eds.) *Qualitative research in practice: Examples for discussion and analysis* (pp. 18-36). San Francisco, CA: Jossey-Bass.
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Second Edition. Thousand Oaks, CA: Sage Publications.
- Moats, L. C. (1999). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. Washington, D.C.: American Federation of Teachers.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institution of Child Health and Human Development (NIH) – sponsoring agency. Report #NIH-00-4769. Pub date: 2000-04-00.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Second Edition. Newbury Park: Sage Publications.
- Patton, M. Q. (2002). *Qualitative evaluation and research methods*. Third Edition. Thousand Oaks, CA: Sage Publications.
- Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A., Flay, B. R., Daniels, S., & Cormack, C. (1990). Effects of program implementation on adolescent drug use behavior. *Evaluation Review*, 14, 264-289.
- Rossi, P., & Freeman, H. (1993). *Evaluation: A systematic approach*. 5<sup>th</sup> Edition, Newbury Park, CA: Sage Publications.

- Seale, C. (1999). *The quality of qualitative research: Introducing qualitative methods*. Thousand Oaks, CA: Sage Publications.
- Schaps, E., Moskowitz, J. M., Malvin, J. H., & Schaeffer, G. A. (1986). Evaluation of seven school-based prevention programs: A final report on the Napa Project. *The International Journal of the Addictions, 21*, 1081-1112.
- Senge, P. (1990). *The fifth discipline*. New York: Doubleday.
- Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)*. (Technical Report) Eugene, OR: University of Oregon.
- Simmons, D. C., Kame'enui, E. J., Good, R. H., Harn, B. A., Cole, C., & Braun, D. (2001). Building, implementing, and sustaining a beginning reading improvement model school by school and lessons learned. In M. Shinn, G. Stoner & H. M. Walker (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches*. Washington, DC: National Association of School Psychologists.
- Skiba, R., Wesson, C., & Deno, S. (1982). The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.
- Smith, S. B., Baker, S., & Oudeans, M. K. (2001). Making a difference in the classroom with early literacy instruction. *Teaching Exceptional Children, 33*, 8-14.
- Smith, S. B., Simmons, D. C., & Kame'enui, E. J. (1998). Phonological awareness: Research bases. In D. C. Simmons & Kame'enui, E. J. (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics* (pp. 61-127). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stanovich, K. E. (1986). Mathew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.
- Stanovich, K. E., & Stanovich, P. J. (1995). How research might inform the debate about early reading acquisition. *Journal of Research in Reading, 18*, 87-105.
- Stecker, P., Fuchs, L., & Fuchs, D. (2005). Using Curriculum-based Measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.

- Supovitz, J., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Report, Philadelphia, PA: Consortium for Policy Research and Education,.
- Taylor, S. J. and Bogdan, R. (1998). *Introduction to qualitative research methods*. Third Edition. New York: John Wiley & Sons, Inc.
- Tindal, G., Fuchs, L., Mirkin, P., Christenson, S., & Deno, S.(1981). *The relationship between student achievement and teacher assessment of short- or long-term goals* (Research Report no. 61). Minneapolis: Institute for Research on Learning Disabilities.
- Torgesen, J. K. (2002). *Florida's Reading First assessment plan: An explanation and guide*. Retrieved March 20, 2006, from Florida Center for Reading Research website ([www.fcrr.org/assessment/PDFfiles/Fl\\_Assess\\_Plan\\_Final.pdf](http://www.fcrr.org/assessment/PDFfiles/Fl_Assess_Plan_Final.pdf)).
- U.S. Department of Education (n.d.). *Reading First*. Retrieved March 20, 2006, from [www.ed.gov/programs/readingfirst/index.html](http://www.ed.gov/programs/readingfirst/index.html).
- Wayman, J. C. & Stringfield, S. (2003). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education*, 112, 549-571.
- Wesson, C. (1991). Curriculum-based Measurement and two models of follow-up consultation. *Exceptional Children*, 57, 246-256.
- Wesson, C., Deno, S., Mirkin, P., Sevcik, B., Skiba, R., King, R., Tindal, G., & Maruyama, G. (1988). A causal analysis of the relationship among ongoing measurement and evaluation, structure of instruction, and student achievement. *The Journal of Special Education*, 22, 330-343.
- Wesson, C., King, R., & Deno, S. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? *Learning Disability Quarterly*, 7, 45-48.
- Wesson, C., Skiba, R., Sevcik, B., King, R., Tindal, G., Mirkin, P., & Deno, S. (1983). *The impact of the structure of instruction and the use of technically adequate instructional data on reading improvement*. Research Paper # 116. University of Minnesota: MINN. Institute for Research on Learning Disabilities.
- Wesson, C., Skiba, R., Sevcik, B., King, R., and Deno, S. (1984). The effects of technically adequate instructional data on achievement. *Remedial and Special Education*, 5, 17-22.

Yell, M., Deno, S., Marston, D. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique, 18*, 99-112.

Young, V. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*, 521-548.

## Appendix A

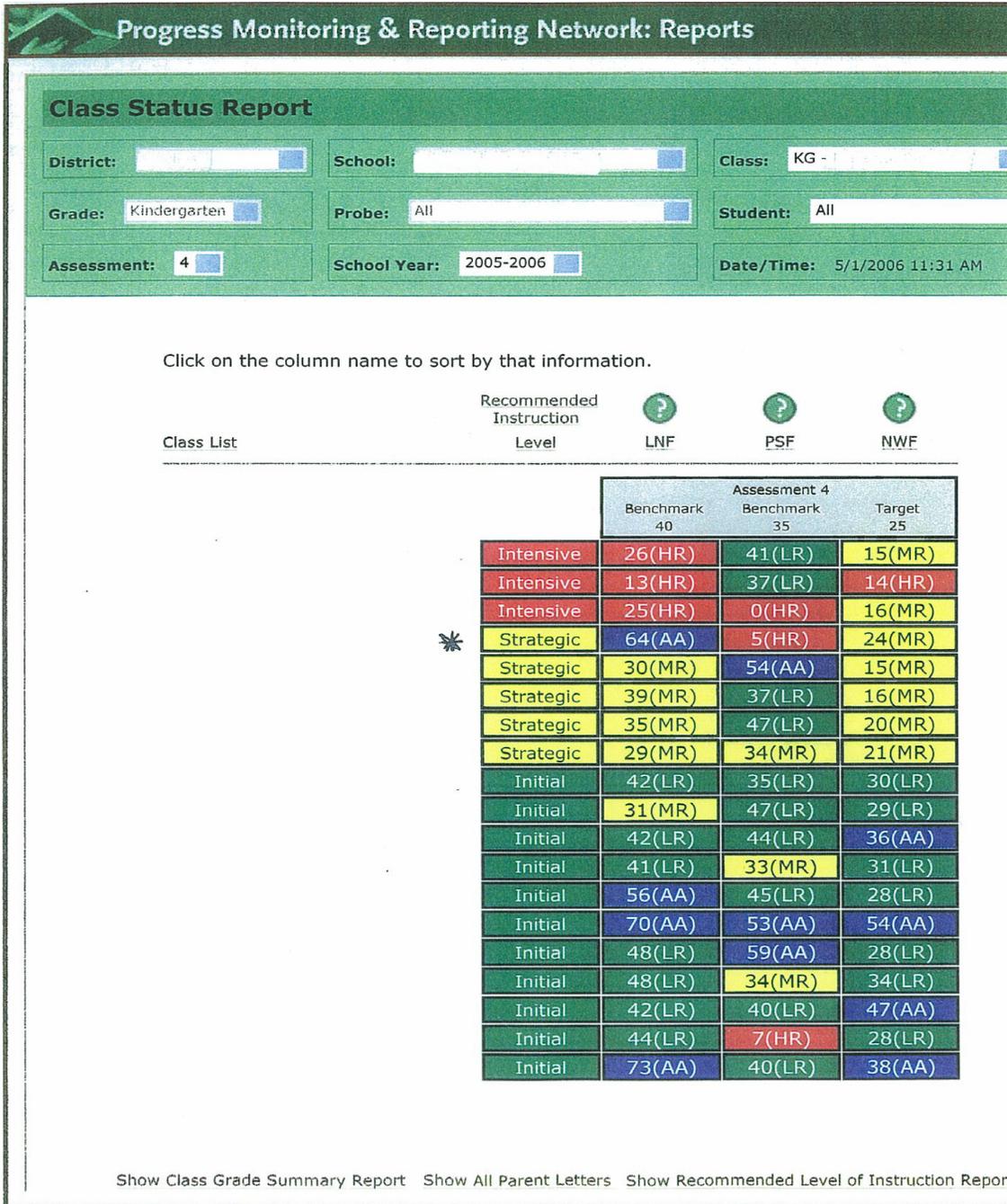


Figure 1: Example of a Class Status Report which is one type of progress report available through the PMRN. This type of report allows for analysis of both class and individual needs. Colors used indicate severity of need for additional supports or instructions to meet benchmark goals. This type of report may be rank ordered by student alphabetical order or student instructional need.

### Student Grade Summary Report

District: <input type="text"/>	School: <input type="text"/>	Class: KG - <input type="text"/>
Grade: Kindergarten <input type="text"/>	Probe: All <input type="text"/>	Student: <input type="text"/>
Assessment: 4 <input type="text"/>	School Year: 2005-2006 <input type="text"/>	Date/Time: 5/1/2006 11:33 AM

#### Kindergarten

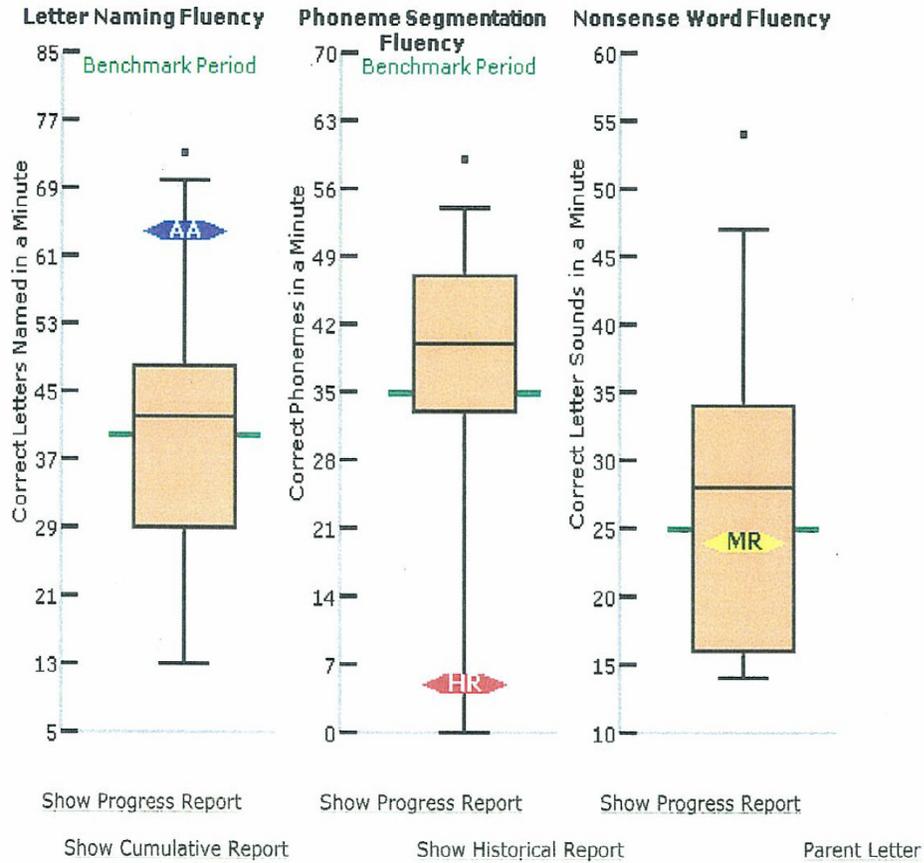


Figure 2: Example of a Student Grade Summary Report which is one type of summary report. This type of report allows for an analysis of individual and classroom needs in relation to benchmark goals for a particular assessment period.

## Reading Progress Monitoring Student Cumulative Report

**Student:** \_\_\_\_\_ **Date:** \_\_\_\_\_ **School Year:** 2005-2006

**Assessment 1** **District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	6	MR	32	AA	-	-	-	-	-	-

**Assessment 2** **District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	8	MR	40	AA	-	-	-	-	-	-

**Assessment 3** **District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	4	HR	45	AA	0	HR	0	HR	-	-

**Assessment 4** **District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	-	-	64	AA	5	HR	24	MR	-	-

### Year End Outcome Tests

PPVT		Stanford-10		GMRT		FCAT		
Score	Rank	Score	Rank	Score	Rank	SSS	Level	NRT
72	19							

### Key

RIL = Recommended Instructional Level  
 iii = Intensive  
 s = Strategic  
 ii = Initial

Risk Levels  
 HR = High Risk  
 MR = Moderate Risk  
 LR = Low Risk  
 AA = Above Average  
 - = Test Not Administered

### Note

Progress Monitoring Instruments  
 K-5 = Dynamic Indicators of Basic Early Literacy Skills (DIBELS)  
 6-8 = Florida Oral Reading Fluency

Figure 3: Example of Reading Progress Monitoring Student Cumulative Report which is one type of a cumulative report offered by the PMRN. This type of report offers an analysis of a student's progress throughout the school year in relation to periodic benchmark goals and end of year goals. It also allows for comparative analysis of alternative assessments provided throughout the school year (e.g., *Peabody Picture Vocabulary Test*)

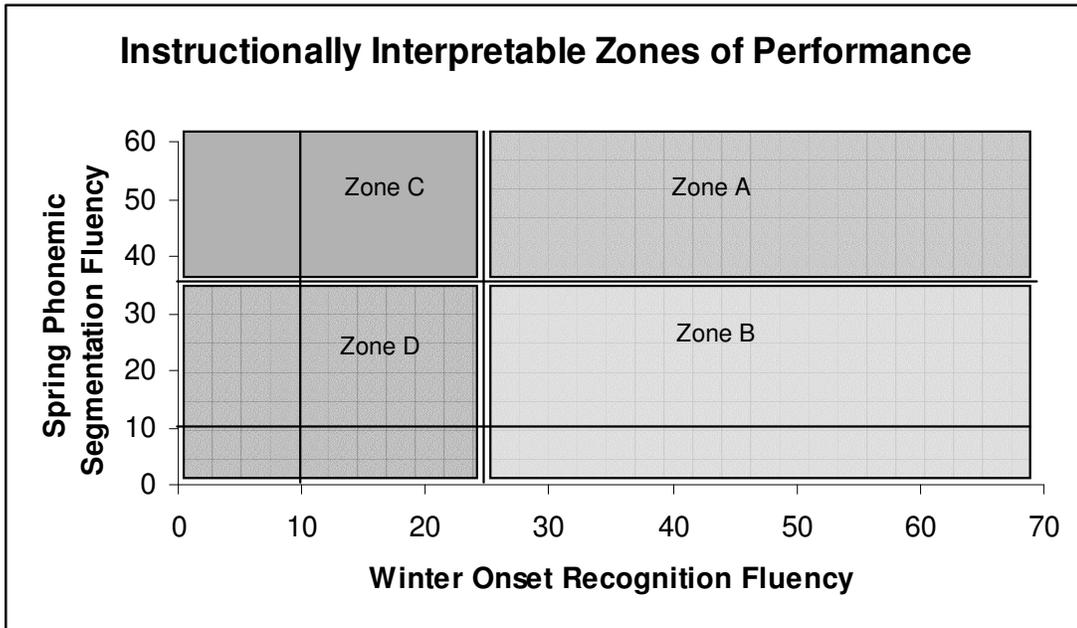


Figure 4: Graph developed by Good et al. (2000) for use in a Benchmark Linkage Report which may help identify a school's core curriculum needs and student instructional needs by analyzing performance on earlier benchmark goals in relation to later benchmark goals.

## Appendix B

Appendix B: Table 3

Number of schools sampled across the three groups of participants.

Schools	Teachers	Reading Coaches	Specialists
1	X		X
2			X
3	X		
4		X	
5	X		
6	X	X	
7	X		X
8		X	
9	X		X
10			X
11		X	X
12		X	
13	X	X	X
14			
15	X	X	
16	X		
Totals Schools Sampled	9	8	7

Appendix B: Table 4  
Probability Sampling Table Developed by DePaulo (2000)

Population Incidence	Probability of Missing a Population Subgroup in a Random Sample							
	Number of Respondents							
	10	20	30	40	50	60	100	200
.50	.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
.33	.018	<.001	<.001	<.001	<.001	<.001	<.001	<.001
.25	.056	.003	<.001	<.001	<.001	<.001	<.001	<.001
.20	.107	.012	.001	<.001	<.001	<.001	<.001	<.001
.10	.349	.122	.042	.015	.005	.002	<.001	<.001
.05	.599	.358	.215	.129	.077	.046	.006	<.001
.01	.904	.818	.740	.669	.605	.547	.366	.134

Appendix B: Table 5

Microsoft Excel format used for organizing, coding, and sorting transcript data

---

Topic #	Participant ID	Data Entry #	Data Segment
1300	FA	1	Text Text Text
2300	FA	21	Text Text Text

Appendix B: Table 6

Comparison of themes identified across multiple sources (teachers, Reading Coaches, and specialists) to organize findings for presentation.

<u>Data Source</u>	<u>Climate/Culture</u>	<u>Resources/Supports</u>	<u>Knowledge of DIBELS</u>	<u>Collecting and Using data</u>
<b>Teachers</b>				
Benefits of DIBELS	x	x	x	x
Concerns of DIBELS	x	x	x	x
General Comments of Impact of Reading First Grant/DIBELS	x	x		
Knowledge of DIBELS			x	
Collecting Data				x
DIBELS vs. Other Assessments			x	x
Issues Related to Monitoring Student Progress		x		x
General Comments about Assessments	x			x
Using Data		x		x
Usefulness/Benefits to PMRN reports		x	x	x
Student reactions to DIBELS				x
Teacher Self-Efficacy	x		x	x
Support for teachers to use DIBELS		x		
Expectations/Emphasis/Pressures	x			
Advice for using DIBELS	x	x		
<b>Reading Coaches</b>				
Role of Reading Coach		x	x	x
Reading Coaches Perceptions of Teachers using DIBELS/PMRN	x	x	x	x
RC perceptions of factors influencing teacher use of DIBELS	x	x	x	x
Student's Reactions to DIBELS				x
RC's concerns about teachers/schools using DIBELS		x	x	x
Advice from RC for using DIBELS		x		x
Role of Leadership	x			x
DIBELS vs. Other Assessments			x	x
RC's perceptions of teacher concerns of DIBELS subtests	x	x	x	
<b>Specialists</b>				
Comparisons b/w RF and nonRF schools				x
Role and Importance of RC		x		x
Specialists perceptions of factors influencing teachers' use of DIBELS	x	x		x
Actions used to increase teachers value or use of DIBELS		x	x	x
Progress Monitoring		x		x
Advice for future		x		x
Specialists' perceptions of teachers' value of DIBELS	x		x	x
Data collection procedures				x
Specialists' perceptions of teacher concerns about DIBELS subtests		x	x	
Teachers' ability for data analysis		x		x

Appendix B: Table 7

School district assessment schedule for kindergarten and first grades.

Tests	Frequency
(KG) school readiness test	1 x at start of KG year.
(KG) district assessment	5 x in year (at least) – many use for progress monitoring also
(1 <sup>st</sup> ) Running Record Assessments	5 x in year (at least) – many use for progress monitoring also
(KG/1 <sup>st</sup> ) DIBELS	3 x in year
(KG/1 <sup>st</sup> ) Peabody Picture Vocabulary Test	1 x at end of the year
(1 <sup>st</sup> ) Stanford Achievement Test – 10	1 x at end of the year

## Appendix C

## Teacher Interview Guide

(30 minutes)

**1. As a teacher at a *Reading First* school, can you help me understand what DIBELS is and what it is used for?**

*Follow-Up Questions (if needed)*

- How helpful are DIBELS data for you and why?
- Do you have any concerns about using DIBELS and why?

**2. I understand that all teachers at *Reading First* schools have access to a web-based program called the Progress Monitoring and Reporting Network which allows teachers access to their students' DIBELS data. Do you use this program, and if so, how helpful has it been for you?**

*Follow-Up Questions (if needed)*

- If no, why not? How do you receive your class information?
- If yes:
  - What types of graphs do you use and why?
  - What do you use the graphs for?
  - How often do you log onto this program?
- Do you have any suggestions about how to improve this data-based program?

**3. How often do you collect DIBELS data for students who are struggling in your class in reading?**

*Follow-up Questions (if needed)*

- Do you collect progress monitoring data for students who are struggling?
- If yes: Why do you collect such information?
- If no: How do you track the progress of students who are struggling?
- If teacher does not collect such data, Who collects such information?

**4. Have you received, or participated in, any training in the use of DIBELS; either one-on-one with your Reading Coach, or through formal training with the school district? If so, could you briefly tell me what you learned and how helpful that/those trainings were?**

*Follow-Up Questions (if needed)*

- Can you help understand what the different subtests measure?
- Have you learned what the 5 skill areas are in reading? Would you help me with what they are?
- Do you know which of the 5 skill areas DIBELS is intended to measure?
- What are teachers supposed to do with this data?

**5. What other types of assessments are conducted throughout the school year in your classroom or school, and do you have any concerns about how to use DIBELS in addition to using other assessments?**

*Follow-Up Questions (if needed)*

- How do you obtain the DIBELS data for your class?
- Does anyone help you evaluate the scores and determine what to do with them?
- What is the most frequent use of the DIBELS data for you?

- *Does DIBELS data provide you with any information that other assessments do not? If so, what?*

**6. What supports are available for you in understanding how to interpret your students' scores?**

*Follow-Up Questions (if needed)*

- *What role does your school's Reading Coach have? What does he/she do?*
- *What learning opportunities are there to learn more about using DIBELS?*
- *Would you say that your school strongly supports DIBELS use? Why or Why not?*

**7. In final, do you have any comments or thoughts regarding the use of DIBELS that leaders at the District Level and State Level should know about from a classroom teacher's perspective?**

*Follow-Up Questions (if needed)*

- *If teacher responds with general statement of overwhelming amounts of assessments in district:*
  - *It sounds like there is a lot going on in this school district. Can you help me understand how this is affecting your perceptions about DIBELS?*
  - *Do you feel that DIBELS has been valuable for you? Why or Why not?*
  - *Would you like to receive more training or information about using DIBELS?*

**Thank you so much for you time!**

## Appendix D

## Appendix D1

### Case Study Questions

(30 minutes)

**I have a case study that I would like to get your feedback on. It is for a student who is in kindergarten/the first grade. The student has not repeated any grades. The student is only having difficulties in reading. Could you take a look at this student's data and tell me what your impressions are about this student?**

*Follow-Up Questions (if needed)*

- *What would you do with a student like this?*
- *What do you feel this student needs?*
- *What other types of information would you need to know to help this student?*
- *What difficulty(s) is this student having based on this information?*
- *What interventions would you suggest to help this student?*
- *Have you seen these types of reports before?*
- *Do you prefer to use other types of reports for looking at individual students?*

Appendix D2: Case Study #1: Kindergarten Student Class Status Report

**Progress Monitoring & Reporting Network: Reports**

### Class Status Report

District:  School:  Class: KG -

Grade: Kindergarten  Probe: All  Student: All

Assessment: 4  School Year: 2005-2006  Date/Time: 5/1/2006 11:31 AM

Click on the column name to sort by that information.

Class List

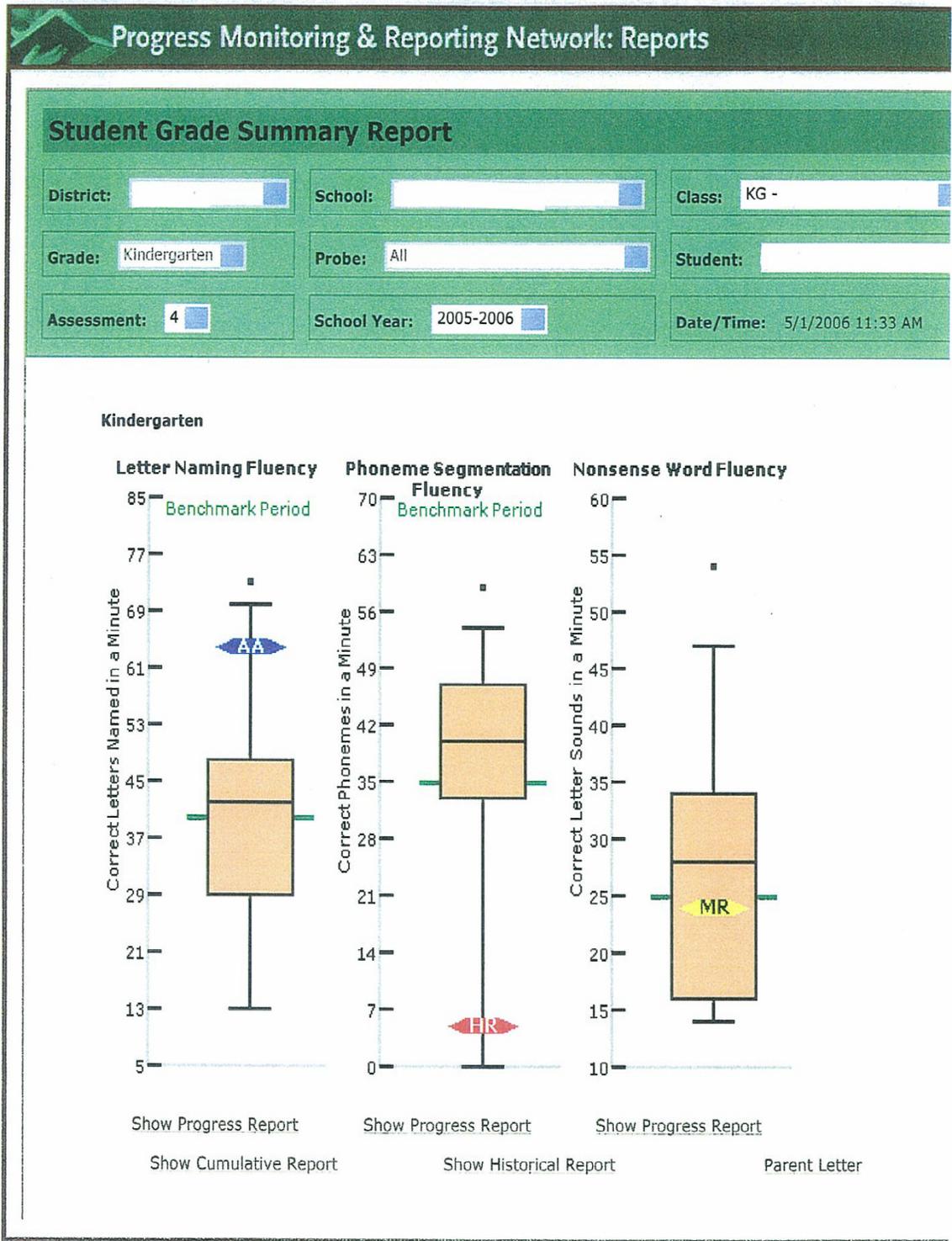
Recommended Instruction Level

LNF PSF NWF

	Assessment 4		
	Benchmark 40	Benchmark 35	Target 25
Intensive	26(HR)	41(LR)	15(MR)
Intensive	13(HR)	37(LR)	14(HR)
Intensive	25(HR)	0(HR)	16(MR)
* Strategic	64(AA)	5(HR)	24(MR)
Strategic	30(MR)	54(AA)	15(MR)
Strategic	39(MR)	37(LR)	16(MR)
Strategic	35(MR)	47(LR)	20(MR)
Strategic	29(MR)	34(MR)	21(MR)
Initial	42(LR)	35(LR)	30(LR)
Initial	31(MR)	47(LR)	29(LR)
Initial	42(LR)	44(LR)	36(AA)
Initial	41(LR)	33(MR)	31(LR)
Initial	56(AA)	45(LR)	28(LR)
Initial	70(AA)	53(AA)	54(AA)
Initial	48(LR)	59(AA)	28(LR)
Initial	48(LR)	34(MR)	34(LR)
Initial	42(LR)	40(LR)	47(AA)
Initial	44(LR)	7(HR)	28(LR)
Initial	73(AA)	40(LR)	38(AA)

Show Class Grade Summary Report Show All Parent Letters Show Recommended Level of Instruction Report

Appendix D2: Case Study #1: Kindergarten Student Grade Summary Report



Appendix D2: Case Study #1: Kindergarten Reading Progress Monitoring Student Cumulative Report

**Reading Progress Monitoring Student Cumulative Report**

**Student:** \_\_\_\_\_ **Date:** \_\_\_\_\_ **School Year:** 2005-2006

**Assessment 1** District: \_\_\_\_\_ School: \_\_\_\_\_ Grade: KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	6	MR	32	AA	-	-	-	-	-	-

**Assessment 2** District: \_\_\_\_\_ School: \_\_\_\_\_ Grade: KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	8	MR	40	AA	-	-	-	-	-	-

**Assessment 3** District: \_\_\_\_\_ School: \_\_\_\_\_ Grade: KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	4	HR	45	AA	0	HR	0	HR	-	-

**Assessment 4** District: \_\_\_\_\_ School: \_\_\_\_\_ Grade: KG

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	-	-	64	AA	5	HR	24	MR	-	-

**Year End Outcome Tests**

PPVT		Stanford-10		GMRT		FCAT		
Score	Rank	Score	Rank	Score	Rank	SSS	Level	NRT
72	19							

**Key**

RIL = Recommended Instructional Level  
 iii = Intensive  
 s = Strategic  
 ii = Initial

Risk Levels  
 HR = High Risk  
 MR = Moderate Risk  
 LR = Low Risk  
 AA = Above Average  
 - = Test Not Administered

**Note**

Progress Monitoring Instruments  
 K-5 = Dynamic Indicators of Basic Early Literacy Skills (DIBELS)  
 6-8 = Florida Oral Reading Fluency

Appendix D2: Case Study #2: First Grade Class Status Report

**Progress Monitoring & Reporting Network: Reports**

### Class Status Report

District:  School:  Class: 01

Grade: 1st Grade Probe: All Student: All

Assessment: 4 School Year: 2005-2006 Date/Time:

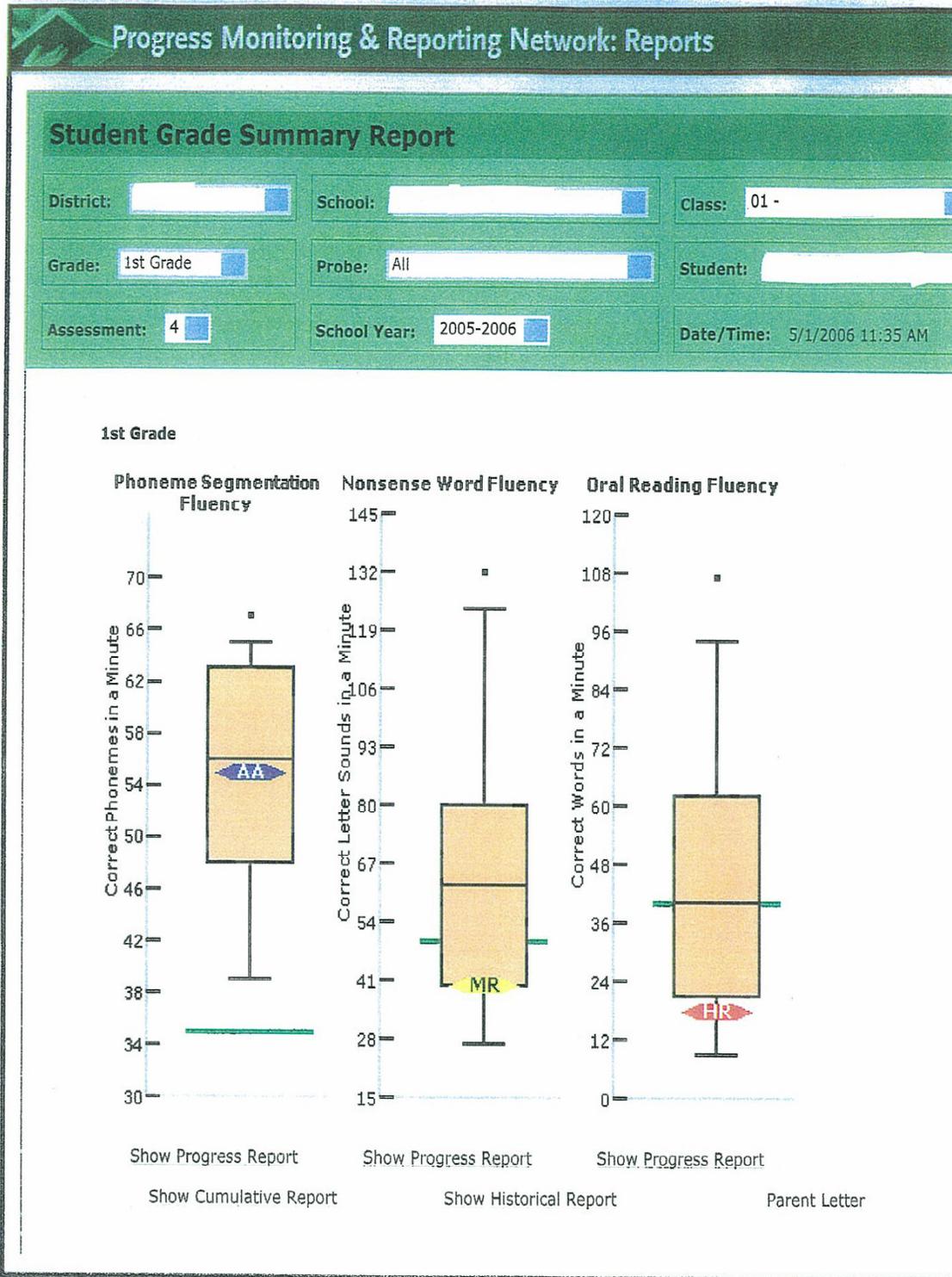
Click on the column name to sort by that information.

Class List Recommended Instruction PSF NWF ORF

	Recommended Instruction Level	Assessment 4		
		Target 35	Target 50	Target 40
*	Intensive	55(AA)	40(MR)	18(HR)
	Intensive	45(LR)	44(MR)	9(HR)
	Intensive	60(AA)	27(HR)	15(HR)
	Strategic	42(LR)	42(MR)	29(MR)
	Strategic	61(AA)	56(LR)	34(MR)
	Strategic	55(AA)	45(MR)	27(MR)
	Strategic	64(AA)	65(LR)	31(MR)
	Strategic	56(AA)	61(LR)	21(MR)
	Strategic	57(AA)	34(MR)	30(MR)
	Initial	63(AA)	132(AA)	88(AA)
	Initial	39(LR)	89(AA)	94(AA)
	Initial	65(AA)	36(MR)	41(LR)
	Initial	48(LR)	80(AA)	66(AA)
	Initial	67(AA)	74(AA)	40(LR)
	Initial	51(LR)	72(AA)	62(LR)
	Initial	50(LR)	116(AA)	107(AA)
	Initial	56(AA)	75(AA)	57(LR)
	Initial	59(AA)	124(AA)	44(LR)
	Initial	65(AA)	62(LR)	41(LR)

Show Class Grade Summary Report Show All Parent Letters Show Recommended Level of Instruction Report

Appendix D2: Case Study #2: First Grade Student Grade Summary Report



Appendix D2: Case Study#2: First Grade Reading Progress Monitoring Student Cumulative Report

**Reading Progress Monitoring Student Cumulative Report**

**Student:** \_\_\_\_\_ **Date:** \_\_\_\_\_ **School Year:** 2005-2006

**Assessment 1 District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** 1st

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
ii	-	-	41	LR	40	LR	29	LR	6	MR

**Assessment 2 District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** 1st

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
ii	-	-	-	-	50	AA	36	MR	14	LR

**Assessment 3 District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** 1st

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
s	-	-	-	-	45	LR	36	MR	14	MR

**Assessment 4 District:** \_\_\_\_\_ **School:** \_\_\_\_\_ **Grade:** 1st

RIL	ISF		LNF		PSF		NWF		ORF	
	Score	Risk								
iii	-	-	-	-	55	AA	40	MR	18	HR

**Year End Outcome Tests**

PPVT		Stanford-10		GMRT		FCAT		
Score	Rank	Score	Rank	Score	Rank	SSS	Level	NRT
83	13	27	34					

**Key**

RIL = Recommended Instructional Level  
 iii = Intensive  
 s = Strategic  
 ii = Initial

Risk Levels  
 HR = High Risk  
 MR = Moderate Risk  
 LR = Low Risk  
 AA = Above Average  
 - = Test Not Administered

**Note**

Progress Monitoring Instruments  
 K-5 = Dynamic Indicators of Basic Early Literacy Skills (DIBELS)  
 6-8 = Florida Oral Reading Fluency

## Appendix E

## Focus Group Guide

(1 hour)

1. **As a *Reading First* school you are required to collect DIBELS data at least 4 times per year. What are your thoughts concerning the use of DIBELS at your school? (10 minutes)**
  - *How do you feel about the use of DIBELS?*
  - *What changes have occurred at your school since DIBELS has been in use?*
  - *Has DIBELS been helpful in improving student achievement at your school?*
  - *Do you have any concerns regarding its use at your school?*
2. **What are your thoughts or feelings concerning teacher's use of DIBELS? (10 minutes)**
  - *How do teachers use the data?*
  - *What problems or benefits arise with teacher's use of DIBELS?*
  - *What is your opinion about how to support teachers' use of DIBELS data?*
3. **How are DIBELS data being utilized? (10 minutes)**
  - *What challenges does your school face regarding the use of DIBELS data?*
  - *Have any benefits occurred by using DIBELS data?*
  - *How do you feel about using the Progress Monitoring and Reporting Network – (the web-based reports offered by the Florida Center for Reading Research).*
  - *How do teachers feel about using the reports offered by FCRR?*
  - *What reports do you feel offer the most information or the most useful information? Why?*
4. **How does DIBELS compare to other assessments provided in your school? (10 minutes)**
  - *Do you feel DIBELS provides the necessary information your school needs?*
  - *What is your opinion regarding the use of multiple assessments in reading at your school?*
  - *What assessments are most helpful for screening and monitoring student achievement in reading in your opinion and why?*
5. **What additional comments, concerns, or suggestions do you all have about using DIBELS at your school that you feel is important for others to know? (10 minutes)**
  - *Are there any changes that you would like to see happen at either the district or state level with regards to using DIBELS?*
  - *Has the use of DIBELS been helpful in your school meeting its goals for students?*
  - *Are there any concerns or problems that you would like district or state leaders to address with regards to using DIBELS?*

**Thank you All for you time!!!**

## Appendix F

Appendix F1

Field Notes  
Teacher Interviews  
Participant Code: \_\_\_\_\_

Date Collected: \_\_\_\_\_

Date Summarized: \_\_\_\_\_

1. Gender: Male \_\_\_\_ Female \_\_\_\_
2. Current Grade Teaching \_\_\_\_
3. Age: \_\_\_\_ 21-25      \_\_\_\_ 26-30      \_\_\_\_ 31-35      \_\_\_\_ 36-40      \_\_\_\_ 41-45  
          \_\_\_\_ 46-50      \_\_\_\_ 51-55      \_\_\_\_ 56-60      \_\_\_\_ 61-65      \_\_\_\_ 66+
4. Ethnicity:  
    \_\_\_\_ Caucasian  
    \_\_\_\_ African American  
    \_\_\_\_ Hispanic/Latino  
    \_\_\_\_ Asian/Pacific Islander  
    \_\_\_\_ Native American /Alaska Native  
    \_\_\_\_ Other (specify): \_\_\_\_\_
5. Number of years experience teaching reading at current grade level?  
    \_\_ <1    \_\_ 1-5    \_\_ 6-10    \_\_ 11-15    \_\_ 16-20    \_\_ 21-25    \_\_ 26+
6. Number of years teaching at present school in the current grade level being taught?  
    \_\_ <1    \_\_ 1-5    \_\_ 6-10    \_\_ 11-15    \_\_ 16-20    \_\_ 21-25    \_\_ 26+
7. Highest Degree Attained:  
    \_\_ Bachelors    \_\_ Masters    \_\_ Specialist    \_\_ Doctorate    \_\_  
    Other: \_\_\_\_\_
8. Credentials (Teacher Certification Areas) – Please mark all that apply:  
    \_\_ K-12 General Education Teacher  
    \_\_ Special Education Teacher  
    \_\_ Reading Endorsement  
    \_\_ Principal/Administrator  
    \_\_ Other (specify): \_\_\_\_\_
9. Average number of hours of participation in professional development activities in reading per year?  
    \_\_ 0    \_\_ 1-5    \_\_ 6-10    \_\_ 11-15    \_\_ 16-20    \_\_ 21-25    \_\_ 26-30    \_\_  
    30+
10. Any previous experience teaching students with Learning Disabilities as an ESE teacher:  
    \_\_ YES: If so, how many years experience: \_\_\_\_\_  
    \_\_ NO

11. Is your Reading Block scheduled for 90 minutes per day?

\_\_\_ YES

\_\_\_ NO: If not, how many minutes per day? (specify): \_\_\_\_\_

12. How many students do you currently have in your classroom? (specify): \_\_\_\_\_

13. Descriptions of the classroom environment (physical arrangement):

---

---

---

---

---

---

---

---

---

---

14. Other relevant notes to include concerning physical arrangement and/or instructional practices observed:

---

---

---

---

---

---

---

---

---

---

15. Notes concerning interview – ideas, thoughts, feelings concerning the interview:

---

---

---

---

---

---

---

---

---

---

## Appendix G

## Appendix G1 – Teacher Interview Codes Draft 1

### Interview Codes:

- 100 – Description of DIBELS
- 200 – Use of Colors on DIBELS
- 300 – General perceptions of PMRN reports
- 400 – Specific PMRN reports preferred
- 500 – Using PMRN reports to inform instruction
- 600 – Understanding of how fluency relates to comprehension
- 700 – Using DIBELS for progress monitoring
- 800 – How to monitor students' progress if not DIBELS
- 900 – What data are used to inform instruction
- 1000 – Participation in DIBELS training
- 1100 – PMRN Training
- 1200 – DIBELS vs. Other Assessments (comparisons)
- 1300 – Accessing the PMRN (Technical skills)
- 1400 – Teacher's perceptions of school culture regarding DIBELS
- 1500 – Using nonsense words
- 1600 – Perceptions of the Nonsense Word Fluency subtest.
- 1700 – Good readers who don't comprehend what they read
- 1800 – Conflict between reading nonsense words and class strategies for unknown words
- 1900 – Instructional strategies for reading unknown words
- 2000 – Setting the climate for students taking the DIBELS
- 2100 – What is needed to maintain use of DIBELS into future at school
- 2200 – What state and district leaders should know from a teacher's perspective
- 2300 – How DIBELS is used to assign students to interventions/services
- 2400 – Literacy Success Program/Title I services
- 2500 – Administrator support for DIBELS
- 2600 – Concerns about using DIBELS
- 2700 – What teachers like about DIBELS
- 2800 – Do teachers think DIBELS is valuable
- 2900 – What teacher's think the DIBELS measures
- 3000 – What teacher's think the Nonsense Word Fluency measures
- 3100 – Do teacher's need more training on DIBELS
- 3200 – What additional training do teachers want

## **DIBELS**

### **D1 Knowledge of DIBELS Test/Measures**

Description of DIBELS testing cycles  
What teacher's think the DIBELS measures  
What teacher's think a particular subtest measures  
Teachers' knowledge of the subtests used for their grade level  
In first year, teachers knew little about the DIBELS and were generally skeptical

### **D2 Procedures for collecting and distributing results of DIBELS**

School procedures for providing data to teachers  
Timeliness of procedures to collect and distribute DIBELS data

### **D3 DIBELS Training**

Participation in DIBELS training  
Specific requests for additional training  
Teachers feel they don't have sufficient training on what the DIBELS does  
Need to know why it is important as a tool in relation to teaching

### **D4 Using the DIBELS to monitor students and Make Decisions**

Using DIBELS for progress monitoring  
How DIBELS is used to assign students to interventions/services  
Teachers' ability/motivation to interpret DIBELS data  
Grade level data used to identify students in need of help (e.g., Great Leaps)  
Staffing children for ESE consideration using the DIBELS  
Making retention decisions using DIBELS data  
Three cycles of data collection are not enough  
Not enough emphasis is placed on student progress rather than on meeting standards

### **D5 Teacher Observations of Students' reactions to DIBELS**

Setting the climate for students taking the DIBELS  
Concerns about timing children on the DIBELS  
Children frustrated with being tested by a stranger or different location  
Children who are still below benchmarks but making progress need lots of praise and continued encouragement

### **D6 Teacher Perceptions of School Support for DIBELS**

Teacher's perceptions of school culture regarding DIBELS  
What is needed to maintain use of DIBELS into future at school

Concerns that there is an overemphasis on the use of DIBELS  
DIBELS viewed negatively b/c of association with politics at school  
More collaborative environment after first couple of years

**D7 Teachers' Perceptions of the Value of DIBELS**

DIBELS vs. Other Assessments (comparisons)  
Concerns about using DIBELS with children who have speech problems  
Concerns about teacher competitiveness over DIBELS results  
What teachers like about DIBELS  
Value of the Nonsense Word Fluency subtest  
Lack of agreement between DIBELS and other assessments  
Concerns about accuracy or consistency of the DIBELS measures over time  
DIBELS takes less time to complete compared to other assessments  
DIBELS interrupts classtime  
Teachers feel they need to teach to the test since it is timed  
Teachers don't value it if they don't understand it's usefulness  
Concerns with interpreting quantitative information w/o qualitative information  
Teachers feel it is a depressing test because the standards are so high  
Most of the time DIBELS is not helpful because observations in small groups/classroom assessments already indicate what students need help  
Teachers like that someone else does the DIBELS assessment

**PMRN**

**P1 Teachers' perceptions of the value of PMRN Reports**

Concerns/Problems with using PMRN reports  
What teachers like about using the reports  
General perceptions of PMRN reports  
Results should be available faster/Reports take too long  
Using the Parent Letter report to communicate with parents  
Concerns that the parent report does include "above average" praise types of feedback.  
Use Parent Letter to solicit support from parents and offer ideas for use at home  
Parent report too much for some parents  
Add notes to parent report for parents to read, or highlight important parts of the report  
Use "box and whisker" formats to compare classes and school to other schools

**P2 Technical Skills(UNDERSTANDING) How to Access and Read PMRN Reports**

Use of Colors on PMRN  
PMRN Training  
Accessing the PMRN (Technical skills)  
Teachers need constant reward/encouragement to use the reports/PMRN website

**P3 Using PMRN Reports to Identify Student Needs**

Specific PMRN reports preferred  
Using PMRN reports to inform instruction

## **READING COACHES**

**RC1 Teachers' Perceptions of the Value of Reading Coaches**

Concerns about Reading Coaches  
Likes about Reading Coaches  
RC is always available for help if needed

**RC2 What Reading Coaches do in the Schools**

RC provides training and support  
RC assists with data analysis with DIBELS.  
RC provides assistance with intervention and instructional ideas  
RC assists with how to make sense of conflicting data from multiple data sources  
Takes teachers to other schools to observe best practices  
Model lessons in the classroom  
Provide books/chapters for teachers to read as professional development

## **READING DEVELOPMENT RESEARCH**

**RD1 Relationship Between DIBELS and Reading Research**

Understanding of how fluency relates to comprehension  
Understanding how DIBELS is directly tied to research on reading development

## **READING INSTRUCTION/INTERVENTIONS**

**RI1 Specific Strategies/Lessons Taught in the Classroom**

Instructional strategies for reading unknown words  
Practice blending sounds  
Practice making rhyming words  
Practice reading nonsense words  
Using nonsense words  
Stretching out words and sounding them out

Encourage students to sound out words  
Model for students how to read with expression/fluency  
Earobics Computer program  
SRA Open Court  
Harcourt Interventions  
Great Leaps

**RI2 Instructional Services in Schools Provided by others**

Literacy Success Program/Title I services

**RI3 Issues with providing interventions in the class**

Not enough time to find or develop individual lessons/interventions for students who are not doing well.

Teachers feel they need more supports or people to help deliver interventions

Behavior problems interfere with teaching and providing instructions

Concerns with limited support at home for students

**TEACHER ABILITY TO INTERPRET AND USE ASSESSMENT DATA**

**T1 Knowledge of how to use multiple sources of data together**

Dealing with conflicting assessment information on a student

Identifying converging information about a student's abilities

Good readers who don't comprehend what they read

**T2 Knowledge or Training on RTI**

RTI will be used at school for next year

**ADMINISTRATOR INVOLVEMENT**

**AI1 Teachers' Perceptions of Value of Administrator Involvement**

Administrator support for DIBELS

Changes in value or importance of DIBELS among administrators

How DIBELS data are used at the building level (by administrators)

Concerns about pressure to perform as a teacher

Feeling that job is on the line

Data is used as means to judge teacher competence

Too much emphasis as an outcomes measure instead of a progress measure

Climate of negativity or positiveness established by the school leadership

DIBELS can become associated with negative politics at a school

Fear of being able to ask for help out of being judged negatively.

Feeling isolated because climate at school prevents asking for help from colleagues  
Need more collaboration among staff

## **ALTERNATIVE ASSESSMENTS**

### **AS1 Using Assessments Other than DIBELS**

How to Monitor Students' Progress if not DIBELS  
What Data are Used to Inform Instruction if Not DIBELS  
Using Running Records to decide placement of groups and decisions  
Description of the PIAP/Kaplan Assessment

## **ADVICE TO EDUCATION LEADERS FROM TEACHERS**

### **AE1 Recommendations to State/District Leaders about DIBELS**

Suggestions about how to revise the DIBELS  
Concerns about using DIBELS  
Additional training/information requested by teachers  
Results need to be more immediately available for use.  
Need more support on how to use the DIBELS information to help students

### **AE2 Statements About Assessments in Schools in General**

What state and district leaders should know from a teacher's perspective  
What additional training do teachers want

## Appendix G3 – Teacher Interview Codes Draft 3

### **CODES**

*There are 8 broad categories of topics listed below. The **BOLD** and **UNDERLINED** topic headings are the codes to use with the transcripts. Across the broad categories there are 28 Topic Headings for use with the transcripts. Sub-headings listed below the bold/underlined headings are merely there to provide some examples of things that would fall under that category.*

#### **1000 - Overall Value of DIBELS**

##### **1100 - Benefits Of DIBELS**

- 1110 Develop and adjust instructional groups
- 1111 Identify specific skills for targeted instruction
- 1112 Helps to monitor student gains and needs
- 1113 Requires more understanding of reading development
- 1114 Allows for differentiated instruction to support all students
- 1115 Access to raw data allows for developing interventions
- 1116 Provides information for supporting student into next grade
- 1117 Corresponds with other assessments and class observations
- 1118 Seems more reliable/objective with someone else assessing the student
- 1119 Use of DIBELS/RF grant has led to increased reading outcomes
- 1120 Instructional ideas learned through DIBELS/RF has proven to be useful in math and writing (e.g., use of differentiated instruction)
- 1121 Provides information about who is entering grade with necessary prerequisite skills

##### **1200 - Concerns About DIBELS**

- 1210 Too much emphasis on fluency and none on comprehension
- 1211 NWF subtest
- 1212 Questionable validity due to:
  - being a timed test
  - doesn't correspond with other assessments or observations
  - students tested in unknown or different settings than classroom
- 1213 Doesn't provide enough information for student going to next grade
- 1214 Too much emphasis on the score when it's only a snapshot
- 1215 Questionable reliability due to:
  - when someone other than teacher is assessing (unfamiliar with student)
  - errors made by people doing the assessment
  - when student is shy or unsure of stranger doing the assessment
  - results inconsistent with other assessments
- 1216 Encourages teaching to the test
- 1217 Not helpful for improving instruction
- 1218 ORF standards are too high
- 1219 Creates negative competition and judgment among teachers

1220 Too little recognition of student gains when they are still below benchmark

**1300 - General Comments of Impact of DIBELS/Reading First Grant**

- 1310 Teachers have mixed views – basically just another test to give.  
1311 Teacher may not know enough about it to appreciate the data  
1312 Younger generation teachers seem more open to it than older generation  
1313 Overall perceptions seem to increase positively with each year  
1314 First impression with DIBELS negatively impacted due to training  
1315 *Reading First* grant has helped to create more team work  
1316 *Reading First* grant has brought many resources that have been very valuable and effective for supporting student learning in reading.  
1317 Like that there is only three cycles than four – too much testing  
1318 Generally ok with DIBELS, but it's not critical to driving instruction  
1319 Little value because it offers little more than already known by teacher who work with them.

**2000 - Conducting Assessments**

**2100 - Knowledge of DIBELS**

- 2110 How often given  
2111 Description as a timed test  
2112 Required assessment  
2113 Knowledge about why timing is important  
2114 Description of subtests

**2200 - Collecting Data - procedures/description**

- 2210 DIBELS – Whole Class all at once approach  
2211 DIBELS – One student at a time approach  
2212 District Assessments – One-on-one  
2213 Comparison of time efficiency to complete different tests

**2300 - DIBELS vs. Other Assessments (Value Comparisons)**

- 2110 Preference of specific test for use in planning instruction
- Influenced by immediacy of feedback to teacher on how to support student
  - Influenced by familiarity of tests available for use
- 2311 Ability to use DIBELS in conjunction with other assessments
- Influenced by knowledge of DIBELS and what it offers
  - Influenced by knowledge of interpret DIBELS data

- Influenced by correspondence of successful outcome between DIBELS and other assessments

- 2312 Use DIBELS only because it is required
- 2313 Absence of district assessment in reading for first grade
- 2314 DIBELS more time efficient and less impacting upon instructional time
- 2315 More emphasis given to other assessments because they drive decisions regarding retention at end of year

**2400 - Issues related to monitoring student progress**

- 2410 Influenced by knowledge of administrating and accessing materials
- 2411 Awareness of option to give DIBELS more often than 3X a year
- 2412 Influenced by level of training to administer and accessibility for use
- 2413 Direct observations during group/class instruction given higher value
- 2414 Influenced by availability of others to give DIBELS
- 2415 Lack of possibility or awareness of option to retest if measurement error suspected – devalues DIBELS.
- 2416 Using or making graphs to chart progress
- 2417 Involving students in monitoring progress and goal setting

**2500 - General Comments about Assessments**

- 2510 Too much testing in schools and it interferes with instruction
- 2511 Too much testing that does not lead to anything valuable to use
- 2512 KG PIAP and 1<sup>st</sup> grade Running Record tests take too long to complete whole class
- 2513 Following tests are given between KG and 1<sup>st</sup> grades collectively:
- KG Flickers (1x at beginning of year)
  - KG PIAP (district assessment) – (5x in year)
  - KG DIBELS (3x in year)
  - KG Peabody Picture Vocabulary Test (1x at end of year)
  - 1<sup>st</sup> Grade Running Record (5x in year)
  - 1<sup>st</sup> Grade DIBELS (3x in year)
  - 1<sup>st</sup> Grade Stanford Achievement Test – 10 (1x at end of year)

**3000 - Using Data**

**3100 - Assigning resources based on DIBELS data (e.g., personnel)**

**3200 - Using DIBELS to determine what interventions to give**

- 3300 - Using DIBELS for placement decisions (e.g., Spec. Ed or Retention)**
- 3400 - Using DIBELS data to determine level or focus of instruction (i.e., individual, small group, classroom, or grade level).**
- 3500 - Collaborative problem-solving with grade level colleagues (i.e., PLC's)**
- 3600 - Effective use influenced by frequency of data collection and access to raw scores**
- 3700 - Using qualitative information about students to help interpret DIBELS**
- 4000 - Progress Monitoring and Reporting Network (PMRN)
- 4100 - Usefulness/Benefits of PMRN reports**  
 4110 Communicating with parents  
 4111 Use of colors to read reports
- 4300 - Preference for specific reports or mention of reports used**
- 4400 - Problems associated with interpreting or accessing reports**
- 4500 - Recommendations for improving the reports/PMRN system**
- 5000 - Student Involvement or Reactions to DIBELS
- 5100 - Student Involvement or Reactions to DIBELS**
- 6000 - Climate/Culture at School/District Related to Assessments and Standards
- 6100 - Teacher Self-Efficacy**  
 6110 Beliefs about educating students  
 6111 Coping with changes in teaching standards/testing  
 6112 Pressures to teach to the test  
 6113 Taking initiative to learn more about DIBELS
- 6200 - Support for Teachers to Use DIBELS**  
 6210 Reading Coach Support  
 6211 Administration Support  
 6212 Training  
 6213 Assistants helping in the classroom

7000 - Advice to District or State Leaders

7100 - Need more people to assist with testing/instruction

7200 - Need more training on the DIBELS

7300 - Access to more intervention materials to use based on DIBELS data

7400 - Share more information about what is working at other schools

7500 - Minimize or streamline the amount of testing that is happening

8000 - Other/Misc.

8100 - Other/Misc.

## Appendix G4 – Teacher Interview Codes Draft 4

### **CODES**

*There are 8 broad categories of topics listed below. The **BOLD** and **UNDERLINED** topic headings are the codes to use with the transcripts. Across the broad categories there are 16 Topic Headings for use with the transcripts. Sub-headings listed below the bold/underlined headings are merely there to provide some examples of things that would fall under that category.*

#### **1000 - Overall Value of DIBELS**

##### **1100 - Benefits Of DIBELS**

- 1110 Develop and adjust instructional groups
- 1111 Identify specific skills for targeted instruction
- 1112 Helps to monitor student gains and needs
- 1113 Requires more understanding of reading development
- 1114 Allows for differentiated instruction to support all students
- 1115 Access to raw data allows for developing interventions
- 1116 Provides information for supporting student into next grade
- 1118 Seems more reliable/objective with someone else assessing the student
- 1119 Use of DIBELS has led to increased reading outcomes
- 1121 Provides information about who is entering grade with necessary prerequisite skills

##### **1200 - Concerns About DIBELS**

- 1211 Concerned about using the NWF subtest
- 1212 Questionable *validity*
- 1213 Doesn't provide any new information
- 1215 Questionable *reliability*
- 1216 Encourages teaching to the test
- 1217 Not helpful for improving instruction
- 1218 ORF standards are too high
- 1219 Creates negative competition and judgment among teachers
- 1220 Too little recognition of student gains when they are still below benchmark

##### **1300 - General Comments of Impact of DIBELS/Reading First Grant**

- 1310 Teachers have mixed views – basically just another test to give.
- 1311 Teacher may not know enough about it to appreciate the data
- 1312 Younger generation teachers seem more open to it than older generation
- 1313 Overall perceptions seem to increase positively with each year

- 1314 First impression with DIBELS negatively impacted due to training
- 1315 *Reading First* grant has helped to create more team work
- 1316 *Reading First* grant has brought many resources that have been very valuable and effective for supporting student learning in reading.
- 1317 Like that there is only three cycles than four – too much testing
- 1318 Generally ok with DIBELS, but it's not critical to driving instruction
- 1319 Little value because it offers little more than already known by teacher who work with them.
- 1320 Any general statement about liking or disliking but without reasons.
- 1321 Instructional ideas learned through *Reading First* are helpful in other areas

2000 - Conducting Assessments

**2100 - Knowledge of DIBELS**

- 2110 How often given
- 2111 Description as a timed test
- 2112 Required assessment
- 2113 Knowledge about why timing is important
- 2114 Description of subtests (without value or judgement)

**2200 - Collecting Data - procedures/description**

- 2210 DIBELS – Whole Class all at once approach
- 2211 DIBELS – One student at a time approach
- 2212 District Assessments – One-on-one
- 2213 Comparison of time efficiency to complete different tests
- 2214 General procedures or description of how long it takes to test class
- 2215 Teachers not administering the DIBELS

**2300 - DIBELS vs. Other Assessments (Value Comparisons)**

- 2110 Preference of specific test for use in planning instruction
- Influenced by immediacy of feedback to teacher on how to support student
  - Influenced by familiarity of tests available for use
- 2316 Using DIBELS in conjunction with other assessments
- Influenced by knowledge of DIBELS and what it offers
  - Influenced by knowledge of interpret DIBELS data

- Influenced by correspondence of successful outcome between DIBELS and other assessments
- 2317 Use DIBELS only because it is required
- 2318 Absence of district assessment in reading for first grade
- 2319 DIBELS more time efficient and less impacting upon instructional time
- 2320 More emphasis given to other assessments because they drive decisions regarding retention at end of year
- 2321 Direct observations during group/class instruction given higher value

**2400 - Issues related to monitoring student progress**

- 2410 Knowledge of administrating and accessing materials
- 2411 Awareness of option to give DIBELS more often than 3X a year
- 2412 Influenced by level of training to administer and accessibility for use
- 2414 Influenced by availability of others to give DIBELS
- 2415 Lack of possibility or awareness of option to retest if measurement error suspected – devalues DIBELS.
- 2416 Using or making graphs to chart progress
- 2417 Involving students in monitoring progress and goal setting
- 2418 Description of using other tests for monitoring progress
- 2419 Observing as monitoring progress

**2500 - General Comments about Assessments**

- 2510 Too much testing in schools and it interferes with instruction
- 2511 Too much testing that does not lead to anything valuable to use
- 2512 KG PIAP and 1<sup>st</sup> grade Running Record tests take too long to complete whole class
- 2513 Following tests are given between KG and 1<sup>st</sup> grades collectively:
  - KG Flickers (1x at beginning of year)
  - KG PIAP (district assessment) – (5x in year)
  - KG DIBELS (3x in year)
  - KG Peabody Picture Vocabulary Test (1x at end of year)
  - 1<sup>st</sup> Grade Running Record (5x in year)
  - 1<sup>st</sup> Grade DIBELS (3x in year)
  - 1<sup>st</sup> Grade Stanford Achievement Test – 10 (1x at end of year)

3000 - Using Data

**3100 - Using Data (any data for any assessment)**

- 3110 Assigning resources based on DIBELS data (e.g., personnel)
- 3111 Using DIBELS to determine what interventions to give
- 3112 Using DIBELS for placement decisions (e.g., Spec. Ed or Retention)
- 3113 Using DIBELS data to determine level or focus of instruction (i.e., individual, small group, classroom, or grade level).
- 3114 Collaborative problem-solving with grade level colleagues (i.e., PLC's)
- 3115 Effective use influenced by frequency of data collection and access to raw scores
- 3116 Using qualitative information about students to help interpret DIBELS
- 3117 Description of using other types of assessment data

4000 - Progress Monitoring and Reporting Network (PMRN)

**4100 - Usefulness/Benefits of PMRN reports**

- 4110 Communicating with parents
- 4111 Use of colors to read reports
- 4112 Preference for specific reports or mention of reports used
- 4113 Problems associated with interpreting or accessing reports
- 4114 Recommendations for improving the reports/PMRN system

5000 - Student Involvement or Reactions to DIBELS

**5100 - Student Involvement or Reactions to DIBELS**

- 5110 Students feeling anxiety or negativity as a response to being tested by the DIBELS.

6000 - Climate/Culture at School/District Related to Assessments and Standards

**6100 - Teacher Self-Efficacy**

- 6110 Beliefs about educating students
- 6111 Coping with changes in teaching standards/testing
- 6112 Pressures to teach to the test
- 6113 Taking initiative to learn more about DIBELS

**6200 - Support for Teachers to Use DIBELS**

- 6210 Reading Coach Support
- 6211 Administration Support
- 6212 Training
- 6213 Assistants helping in the classroom

**6300 - Expecations/Emphasis/Pressures**

- 6310 Too much emphasis on fluency and none on comprehension
- 6311 Too much emphasis on the score when it's only a snapshot
- 6312 State Standards/School Grades by State

7000 - Advice to District or State Leaders

**7100 - Advice related to using the DIBELS**

- 7110 Need more people to assist with testing/instruction
- 7111 Need more training on the DIBELS
- 7112 Access to more intervention materials to use based on DIBELS data
- 7113 Share more information about what is working at other schools
- 7114 Minimize or streamline the amount of testing that is happening
- 7115 Advise on how to modify the DIBELS test

8000 - Other/Misc.

**8100 - Other/Misc.**

*Use this category ONLY if comment talks about anything OTHER THAN reading, assessments, school climate/pressures, or teacher talking about herself. Also use this category if teacher does not understand the question and is not providing any information.*

## Appendix H

**Table of Contents**

I.	General Information.....	2
	a. Purpose of Study.....	2
	b. Research Questions.....	2
	c. Participants.....	2
	d. Data Collection Overview.....	2
	e. General Overview of Findings.....	2
II.	Researcher’s Topics in Interviews/Focus Groups.....	3
III.	Teacher Perceptions and Understandings of the DIBELS.....	3
	a. Climate/Culture of School.....	3
	b. Supports/Resources Available.....	4
	c. Knowledge of DIBELS.....	4
	d. Collecting and Using Data.....	7
IV.	Perceptions of Reading Coaches.....	10
	a. Climate/Culture of School.....	10
	b. Supports/Resources Available.....	10
	c. Teachers’ Knowledge of DIBELS.....	11
	d. Collecting and Using Data.....	12
	e. Advice.....	13
V.	Perceptions of “Specialists”.....	14
	a. Culture/Climate of Schools.....	14
	b. Support/Resources Available.....	15
	c. Teachers’ Knowledge of DIBELS.....	15
	d. Collecting and Using Data.....	15
VI.	Case Study Comparative Analysis between Teachers and DIBELS Experts.....	17
	a. Expert Review of Kindergarten Case Study.....	17
	b. KG teachers’ review of Kindergarten Case Study.....	18
	c. Expert Review of First Grade Case Study.....	18
	d. First grade teachers’ review of First Grade Case Study.....	19
	e. Expert Opinion on the use of DIBELS/ Data in General at Reading First Schools.....	19

## I. General Introduction

### a. Purpose of Study

- i. Descriptive Qualitative Study
- ii. To understand the perceptions and use of DIBELS by teachers at elementary schools with *Reading First* grant.

### b. Research Questions

- i. What attitudes and perceptions exist among persons other than teachers who participate in the collection, input, and analysis of DIBELS data throughout the school year?
- ii. What are teachers' perceptions and understandings about DIBELS and the PMRN?
- iii. How do teachers' understandings and use of DIBELS data, as presented in the PMRN reports, compare to *Reading First* experts who are provided with the same information?

### c. Participants

#### i. Schools

1. Elementary schools in fourth year of Reading First grant implementation.
2. 15 schools total sampled (teachers, reading coaches, specialists)

#### ii. Teachers

1. KG or 1<sup>st</sup> grade teachers only (both combined represent all 5 subtests of DIBELS).
2. At least 2 years teaching present grade at present school
3. Aggregated Demographics
  - a. 14 teachers (7 KG and 7 First)
  - b. Range of years of experience teaching current grade = 1-26 years; 7 teachers with less than 10 years experience and 7 above 10 years experience.
  - c. Range of age = 21-65.
  - d. Range of credentials – 11 teachers with Bachelor's degree in K-12 Teaching, 3 with Masters degree.

#### iii. Reading Coaches

1. 8 participants

#### iv. "Specialists" (involved in collection and/or use of DIBELS at Reading First schools).

1. 6 participants
2. Student Services
3. ESOL

#### v. DIBELS "experts" (individuals at state level who have expertise in use of PMRN/DIBELS data)

1. 2 participants

2. Individual interviews reviewing both KG and 1<sup>st</sup> grade Case Studies
- d. Data Collection
  - i. All involved audio recording for purposes of transcribing and analyzing
  - ii. Transcribed data coded
  - iii. Patterns observed and reported
- e. General Overview of Findings
  - i. Teacher Perceptions of using DIBELS
    1. Complex and unable to sort on dichotomy of like/dislike only.
    2. Perceptions influenced by multiple variables
    3. DIBELS value depends on variable addressed
    4. Differences and Consensus across grades and within grades
  - ii. Reading Coach Perceptions of using DIBELS
  - iii. Specialists' Perceptions of using DIBELS
  - iv. Teacher/Expert Comparisons of utilizing case study involving PMRN reports.
    1. Class Status Report
    2. Student Grade Summary Report
    3. Reading Progress Monitoring Student Cumulative Report

## II. Researcher's Topics in Interviews/Focus Groups

- a. Guided by knowledge of research literature on DIBELS/Reading First Grant
- b. Guided by ongoing observations of comments given by each consecutive interview/focus group
- c. Overall – Topics addressed through interviews/focus groups
  - i. General likes/dislikes about DIBELS
  - ii. General value or perceived impact of Reading First Grant at school
  - iii. Conducting Assessments at school
  - iv. Using data
  - v. PMRN reports
  - vi. Student reactions to being tested with DIBELS
  - vii. Support for teachers in the use of DIBELS at school
  - viii. Advice offered by teachers concerning use of DIBELS
- d. Perceptions of DIBELS can be organized into four main topic areas:
  - i. Climate/Culture of school
  - ii. Support for using DIBELS
  - iii. Knowledge of the DIBELS
  - iv. Collecting/Using assessment data

## III. Teacher Perceptions and Understandings of the DIBELS (What are teachers' perceptions and understandings about DIBELS and the PMRN?)

- a. Climate/Culture of school
  - i. Pressures on teacher performance
    1. Teachers feel general sense of pressure by general climate of accountability (e.g., NCLB) to increase student performance in education.
    2. Most cited teacher competition as initial impact of using DIBELS in first couple of years.
    3. More collaboration evolved over time which led to most teachers becoming less concerned about being evaluated based on students' DIBELS data.
    4. Emphasis of value implicitly or explicitly stated by administrators/district
      - a. Overall, most reported that DIBELS has a greater value placed on it by the school/district than other assessments.
      - b. Most teachers agreed that administrative support and encouragement of use of DIBELS is very important to influence teacher use of DIBELS.
      - c. Some indicated that administrators are important for deflecting pressures of accountability outside of school building.
    5. When asked about the source of pressures being felt, one kindergarten teacher commented that it existed before DIBELS and that much of the resistance towards DIBELS is less about the test than about "removing the childhood from the child." Too much intensity in kindergarten than ever before. "...yes, they get real strong on the DIBELS but their behavior hasn't gone to a place that is appropriate for society."
    6. Pressures also occur on earlier grades from later grades to be sure students are ready to enter next grade level (e.g., the pressure at 3<sup>rd</sup> grade gets transferred to lower grades all the way to kindergarten).
  - ii. Test Overload
    1. Recognition of all tests given and required frequency of test administration
      - a. School readiness assessment (KG) – 1x at start of year.
      - b. District-wide assessment (KG) – 5x in year.
      - c. Running Records (1<sup>st</sup>) – 5x in year
      - d. DIBELS – (KG/1<sup>st</sup>) – 3x in year
      - e. Peabody Picture Vocabulary Test (PPVT) – (KG/1<sup>st</sup>) – 1x at end of year.
      - f. Stanford Achievement Test 10<sup>th</sup> Edition (SAT-10) – (1<sup>st</sup>) – 1x at end of year.
    2. All teachers voiced concerns about the quantity of testing that is taking place in district.

- a. Some voiced amazement that children are even learning at all with all the missed instructional times due to testing.
- b. Only two of the 14 teachers were interested in learning how to give DIBELS themselves for progress monitoring while all others rejected idea because of too much testing already in their responsibility.
- c. Amount of testing and types of testing were important factors in teacher's perceptions of DIBELS (see below).
- 3. All teachers indicated they needed more personnel to help with collecting assessments in the classroom, including DIBELS (Title 1 teachers are pulled away from instruction during DIBELS cycles to help with collecting DIBELS data).
- b. Supports/Resources available through Reading First grant- All teachers praised the Reading First grant for various supports and or resources it provides schools.
  - i. The most valuable is the full-time Reading Coach
    - 1. Helps coordinate DIBELS data collection and data entry
    - 2. Provides technical assistance to teachers about how to use DIBELS data and other assessment data for making instructional decisions.
    - 3. Helps set up/create/introduce various activities/instructional programs for classrooms
    - 4. Helps set up classroom libraries
    - 5. Provides training on DIBELS/PMRN
    - 6. Provides PMRN reports
      - a. All teachers indicated this as important because they don't have access to color printers are dependent on Reading Coach to provide color reports
      - b. Some print their own reports but then have to color them in because they value the color aspects of the PMRN reports (see below)
    - 7. Provides modeling/coaching for teaching reading
    - 8. Coordinates visits to other schools that are modeling effective practice.
    - 9. Accessible and knowledgeable to answer any questions related to reading or DIBELS.
    - 10. Some Reading Coaches established book clubs to help teachers increase their knowledge of reading development and reading instruction
  - ii. Instructional materials was the next most talked about benefit of the Reading First grant.
    - 1. Teachers value the amount of books that they could supply their classrooms with.
    - 2. Various other instructional aids have been valuable to teachers.

- iii. Less common positive impact reported (only 1 teacher) was that Reading First grant and use of DIBELS has forced educators to learn more about reading development and reading instruction.
- iv. Training
  - 1. None of the teachers reported attending any district level trainings on the use of DIBELS. However, one teacher indicated she did get trained on how to administer the DIBELS.
  - 2. Some teachers who see a positive value in the use of DIBELS felt most teachers don't value the DIBELS because they don't have sufficient information about what it is and how to use the information from it to guide instruction – to them it's just another test.
  - 3. All indicated receiving some training from their Reading Coach on school grounds.
    - a. Some training was focused on utilizing various intervention programs for students.
    - b. Those who reported attending the first mandatory training at the start of the grant complained about the format – the trainers read to them word-for-word from the training manual (scripted training). This apparently has been changed in response to teacher concerns over this training format.
    - c. Most indicated some general training or technical assistance received on how to use the colors and numbers to identify struggling students and allocate resources based on the DIBELS information (e.g., Title 1 assistants, participation in Great Leaps reading program, etc.).
  - 4. Some indicated general interest in participating in future trainings if they provided more information about the development and research on DIBELS and how to link it to instruction.
  - 5. Only two of the 14 teachers indicated an interest in learning to use DIBELS for monitoring progress and administering it themselves. Most declined this out of concern that there is too much testing already.
  - 6. Some teachers reported they had not received any formal training on what the DIBELS is, what it assesses, or why it is timed. And yet, others indicated at least receiving such information more informally through working with the Reading Coach or through learning on own.
  - 7. One teacher reported learning that DIBELS Oral Reading Fluency is a predictor of reading success in later grades.
- c. Knowledge of the DIBELS
  - i. DIBELS as a Benchmark Assessment
    - 1. All teachers indicated knowledge of the DIBELS being administered three times a year.

2. All teachers reported knowledge of it being a required test at schools with a Reading First grant.
- ii. DIBELS as a timed test.
  1. Some teachers do not see value in the DIBELS because it is timed.
    - a. Don't see correspondence with DIBELS and district assessments or classroom observations.
      - i. Ex. Student's who do well on district assessment for letter naming in kindergarten or oral reading on Running Records in first grade, but then do poorly on Letter-Naming Fluency or Oral Reading Fluency, respectively, on the DIBELS.
      - ii. Some indicated DIBELS only corresponds with district assessments for students that are struggling (e.g., low on both DIBELS and district assessments)
    - b. 6 of the 7 kindergarten teachers voiced concerns about the appropriateness of timing kindergarten students.
      - i. Some mostly concerned with timing in the first cycle since kindergarten students are not used to being tested.
      - ii. Others feel it is too much pressure on very young children.
      - iii. These teachers were concerned about student's feeling anxious about being timed.
    - c. Most teachers reported DIBELS as a kind of "speed" test.
      - i. Most concerned that the DIBELS values speed instead of comprehension of what is read.
      - ii. Some reported greater value in students reading slow but with comprehension rather than fast without comprehension.
  2. Those that reported positive perceptions of the DIBELS did not indicate any concerns about the DIBELS being timed and also reported finding correspondence with other assessments and/or teacher observations (in both grade levels).
  3. Couple of teachers reported they prepare students for the timing by "making a game of it"
- iii. DIBELS – What it measures
  1. Some teachers reported having difficulty remembering the specific subtests in the DIBELS and/or what the acronyms stand for.
  2. Some teachers reported seeing a correspondence with what the DIBELS measures and what their classroom district assessments measure.
  3. Of those that do see a correspondence between the DIBELS and their district assessments, some don't see value in DIBELS because it doesn't provide any new information beyond what their district assessments provide or what they observe in the classroom.
  4. Some kindergarten teachers did not value DIBELS because they perceive very little correspondence – only Letter Naming Fluency

- and Phoneme Segmentation Fluency (PSF) as being close to what the district assessment measures.
5. Nonsense Word Fluency (NWF)
    - a. 5 of 14 teachers voiced concerns about using the NWF subtest.
      - i. One referred to it as the “silly” test because it uses nonwords.
      - ii. These teachers saw concerns in students being confused on NWF when they try to apply the strategies they are being taught in the classroom to figure out a word they don’t know by thinking of words they do know that look like it or begin with the same sound/letter – the effect is that they lose time.
      - iii. These teachers generally reported a feeling that the NWF subtest is inappropriate because the goal of reading is to help students read real words.
    - b. 5 of the 14 teachers voiced positive uses regarding the NWF subtest.
      - i. When asked about any concerns using the NWF subtest, these teachers generally reported that it tells you if the student can sound out a word they don’t know.
      - ii. Some of these teachers reported value in the NWF subtest for their students to identify their letter-sound correspondence skills (especially with using short vowel sounds).
      - iii. These teachers reported that the raw protocols on which NWF data is collected is more valuable than the NWF score itself because they can observe it for patterns in student errors to inform instruction
    - c. Remaining 4 teachers did not report any perceptions on the use of NWF subtest.
  6. Phoneme Segmentation Fluency (PSF)
    - a. 2 of the first grade teachers reported concerns about the PSF subtest because they feel it confuses students when they are being taught to blend sounds together to make words and then tested to see how they break words into sounds.
    - b. Some teachers (across both grade levels) reported a positive value in the PSF subtest because it is similar to either what is measured in kindergarten on district assessments or because it helps teacher to determine if student has phonemic awareness skills.
  7. Initial Sound Fluency (ISF)
    - a. Only one teacher reported concern about the ISF subtest.

- b. Specifically, feels it is invalid because of the vocabulary labels given for certain pictures that might be labeled something different by a student.
  - c. This teacher did not have any concerns with the procedures for administration of the ISF or the pictures themselves – only the chosen vocabulary labels assigned to certain pictures (e.g., “grass” vs. “yard”).
- 8. Letter Naming Fluency (LNF)
  - a. No teacher reported any concerns about the LNF subtest.
  - b. However, some kindergarten teachers indicated concerns of the timing of DIBELS as leading to lack of correspondence between letter naming ability on DIBELS vs. district assessments (i.e., meeting expectation on one and not the other).
  - c. None of the teachers indicated any value in this subtest except for a general indicator for early kindergarten to see if they are ready to learn how to read.
- 9. Oral Reading Fluency
  - a. First grade teachers have mixed views on this subtest.
  - b. Because of timing, some see it is a kind of speed reading test.
  - c. Others see it as an indicator of being able to comprehend text.
  - d. Some have trouble seeing correspondence between DIBELS ORF and Running Record performances (i.e., low ORF score and high Running Record score).
    - i. Of these specific reports about lack of correspondence with Running Records, none indicated a student profile where ORF is high and Running Record is low.
    - e. Some first grade teachers see it as having value in determining readiness for second grade.
    - f. One teacher felt ORF criteria at end of first grade was too high – concerned that students are being pushed to read fast instead of taking time to comprehend what they read.
- 10. Comprehension - Some teachers indicated they saw less value in the DIBELS because it does not measure comprehension.
- iv. DIBELS as Progress Monitoring Tool
  - 1. All teachers were asked if they use DIBELS for progress monitoring more frequently than the three times a year for benchmark on students who are struggling.
  - 2. All teachers indicated they do not and instead rely on district assessments and/or observations during small group instruction with students to guide instruction day to day.

3. Only one teacher indicated creating graphs to track student performance and the development of teacher-made materials to assess skills using a one-minute timing.
  4. Two teachers indicated an interest in learning how to use DIBELS for progress monitoring and felt that it would be more useful than DIBELS benchmark data to guide instructional decisions for struggling students.
  5. One teacher who was trained in administering DIBELS reported not being able to use DIBELS for progress monitoring due to lack of materials – only access to those being used for Benchmark Assessments.
  6. No teacher reported any knowledge of being able to access DIBELS materials from websites for use in the classroom to monitor progress of students.
  7. When asked if interested in learning how to administer DIBELS for use as a progress monitoring tool, most indicated no interest due to feeling overwhelmed with current testing responsibilities.
- d. Collecting and Using Data
- i. Collecting Data
    1. Two general procedures reported for collecting Benchmark Assessment data on DIBELS:
      - a. Whole class goes to Media Center and work on computers while each student is assessed by one of 5-7 people collecting DIBELS.
      - b. DIBELS testers come to classroom and take one student at a time to nearby quiet area to test.
      - c. Both reported as being quick and efficient processes.
    2. Students working with testers other than teacher
      - a. All teachers indicated a concern, some greater than others, about students being tested by someone they do not recognize or feel comfortable with.
      - b. Some teachers indicated this is mostly a concern only during first cycle.
      - c. Some teachers indicated students less affected when person giving the assessment, or when Reading Coach, visits students before testing to explain the test and set the climate for participation.
      - d. Most students ok because people testing them are often Title 1 personnel who work at the school daily.
      - e. A few teachers indicated value in someone else collecting DIBELS data – seems to make information more objective and reliable when it matches what teacher is seeing in the classroom.
    3. Quality of notes on test protocols

- a. Some teachers indicated high value in qualitative notes added to test protocols to further aid teachers in the interpretation of scores (e.g., not feeling well, shy, etc.)
- b. Teachers reported different testers lead to different use of qualitative notes – some helpful, some not.
- 4. DIBELS vs. district assessments – Benchmark Assessments
  - a. All teachers reported value in how quick DIBELS is collected compared to district assessments in the classroom.
  - b. All teachers reported value in someone else collecting the data – controlling for strangers as testers – because they are overwhelmed with their own assessment responsibilities.
  - c. Although teachers have a value for someone else collecting the DIBELS data, some indicated concerns about not finding as much usefulness in the data because they are not giving the test – cannot observe qualitative aspects of student performance.
    - i. Access to raw data on protocols improves situation.
    - ii. Inclusion of qualitative notes by tester improves situation.
- 5. DIBELS vs. district assessments – Progress Monitoring
  - a. All teachers indicated using a combination of district assessments and teacher observations in the classroom/small group instructional setting to monitor student progress.
  - b. Influenced by lack of training, knowledge, personnel, and/or time to administer DIBELS for progress monitoring.
- ii. Using data
  - 1. Preference for assessment(s) to inform instructional decision making.
    - a. Of the teachers who reported positive perceptions and value on the use of DIBELS, all indicated a preference for using multiple sources of data to inform or guide instruction. At the very least, some of these teachers used DIBELS data to confirm results of other assessments/observations – when different, they seek to understand why.
    - b. Of the teachers who reported negative perceptions and value on the use of DIBELS, all indicated preference for using district assessments and/or observations to inform or guide instruction. Only use DIBELS mainly because it is required – but DIBELS doesn't offer anything new beyond what district assessments and observations reveal about students.
    - c. Preference for what data is used seems to be influenced by knowledge of DIBELS.
    - d. All teachers reported a greater emphasis placed on the use of DIBELS at their schools through meetings with either Reading Coach, administrators, student services personnel, and/or Title 1 personnel.

2. Data to form and adjust instructional groups
  - a. Teachers who prefer a multiple data source approach reported positive value in DIBELS helping to form instructional groups and assist in adjusting group membership through the year.
    - i. Some teachers used DIBELS data to organize their reading center activities.
    - ii. Some teachers reported positive value in using FCRR binder of instructional activities.
  - b. Teachers who preferred district assessments to guide instruction did not report any positive value in using DIBELS to form instructional groups
    - i. Describe DIBELS as “just another test we have to give.”
    - ii. Describe DIBELS as unhelpful because it doesn’t reveal anything new beyond the district assessment data or observations made in classroom.
3. Data to determine interventions or placement in programs
  - a. Working in grade level teams
    - i. All teachers reported they often meet as a grade level team (at least once a month) to discuss data and observations of students who need more help.
    - ii. All teachers reported meeting with Reading Coach at least after each DIBELS cycle to go through DIBELS data and develop strategies to increase grade level/classroom level scores
  - b. Assigning students to reading programs/Small Group Instruction
    - i. Most teachers reported that they use assessment information to identify students who needs more help in reading.
      1. Only some reported explicitly using the DIBELS to determine what students are provided.
    - ii. Some teachers reported specific reading programs that are provided to students who are identified needing help in reading based on either the DIBELS or classroom assessments:
      1. Great Leaps
      2. SRA Open Court
      3. HeadSprout
      4. Teacher developed reading activities
      5. FCRR reading activities (“FCRR binder”)
    - iii. All teachers reported they provide interventions and supports to struggling student during small group instruction
  - c. Title 1/Classroom Volunteers

- i. On the topic of using data, many of the reports given by teachers indicated the use of Title 1 teaching assistants or classroom volunteers.
    - ii. Some reported DIBELS is used to decide which classroom Title 1 support is given.
    - iii. Other reported using DIBELS/classroom assessments to identify who should get Title 1 help.
    - iv. Title 1 small group setting reported as an intervention – only some teachers explicitly indicated use of a program/activity in that setting (e.g., Great Leaps).
  - d. Retention/ESE Referral
    - i. Most of the kindergarten teachers reported using DIBELS/Classroom assessments to make retention decisions and/or to identify students for referral for ESE evaluation/consideration.
    - ii. None of the first grade teachers indicated any use of DIBELS at their school for making retention decisions. One first grade teacher indicated retention decisions were made only based on district assessments.
    - iii. Some teachers (all grades) indicated that DIBELS is looked at by support staff (e.g., student services personnel) when a student is being considered for ESE placement.
- 4. Using PMRN
  - a. Accessing PMRN online
    - i. Most of the teachers (all grades) reported they access the PMRN to download data reports.
      - 1. All of these teachers only use the PMRN for accessing the Parent Letter.
        - a. Parent letter seen as very valuable tool for communicating with parents on where student is instructionally at and how to support at home.
        - b. Some teachers supplement or highlight sections of the Parent Letter to make more useful or efficient for parents (e.g., notes/comments).
      - 2. A few teachers indicated accessing the PMRN more than 3x a year to identify and compare student performances across the year.
    - ii. A few teachers indicated they never access the reports online because they are provided to them by the Reading Coach.
    - iii. Teachers don't have access to color printers.
    - iv. Most of the teachers who access the PMRN or use the reports it generates reported only knowing how to use the colors and scores to identify student needs.

- b. Reading Coach support
  - i. Reading Coach provides reports and technical assistance to grade level teams and individual teachers on interpreting reports at the grade and classroom level.
  - ii. Teachers all indicated, regardless of comfort level for PMRN/Reports or knowledge of DIBELS, that they would not be able to utilize the data as well as they do without the guidance and support of the Reading Coach.
  - iii. All teachers commented great concern about possibly losing their Reading Coach when the grant is expired and the impact it would have on their use of the DIBELS data at their school.
- c. Preference for specific reports (other than parent letter).
  - i. Most showed or described using the Class Status Report as their preference.
  - ii. Some described a “box and whiskers” format.
    - 1. All but one of these teachers reported confusion about that style of data display.
    - 2. One teacher reported learning just recently from the Reading Coach that goal is to get the boxes smaller and above the expectation for the cycle.
  - iii. No other reports were mentioned or reported by teachers.

**IV. Perceptions of DIBELS by Reading Coaches (What attitudes and perceptions exist among persons other than teachers who participate in the collection, input, and analysis of DIBELS data throughout the school year?)**

- a. Climate/Culture of School
  - i. Work load
    - 1. Perceptions there is not enough time to use the assessment or data
    - 2. Teachers have too much on their plate regarding the assessment of students and paperwork – DIBELS as one more thing imposed on them to do.
    - 3. Teachers are burdened with so much paperwork and responsibilities to complete many other types of assessments which offer less information for instructional planning.
    - 4. Teachers are resistant to DIBELS many times because it is one more thing that when added all up, a great deal of instructional time is being lost.
  - ii. Level of administrator involvement, direction, and emphasis on use of DIBELS

1. If administrators do not value data analysis and data utilization along with holding staff accountable, then DIBELS is less likely to be used or valued.
  2. Leadership is essential for the use of DIBELS.
  3. The issue around why some leaders may not demonstrate explicit support for DIBELS may be less to do with their perceived value of it as much as competing demands of other responsibilities placed on them by the district – too much on their plates to have direct and consistent involvement.
  4. Reading Coaches see need to take initiative by taking data to the principal and communicating often with them. Reading Coaches job to show and demonstrate, consistently, the utility of DIBELS data.
  5. Some principals may not have time to review the data often, but entrust a few to do that and provide input to staff.
- iii. Need culture of valuing the use of data to make changes to curriculum/instruction
1. Reading Coaches reported that schools need to have a culture of valuing the use of data to make decisions.
  2. Teachers and administrators many times need to be lead to using the data
- b. Supports/Resources Available
- i. Role of Reading Coach
1. Technical support (data analysis; data utilization; accessing reports)
    - a. Grade level focus
    - b. Classroom level focus
    - c. Individual student level focus
    - d. Identify growth made at each level for each cycle
  2. Modeling lessons/research-based interventions or programs
  3. Ensuring fidelity and integrity of instructional approaches being used
  4. Support teacher needs/questions about instruction and data analysis
  5. On-site training on use of DIBELS
  6. Demonstrate the utility for DIBELS and help others see the value of it
    - a. Meeting with individual teachers
    - b. Meeting with grade level teams
    - c. Meeting with School Leadership Team/Administrators
    - d. Develop plans to increase growth to next cycle
    - e. Share raw data sheets with teachers and encourage them to review them for patterns.
  7. Coordinate DIBELS collection efforts
    - a. Benchmark assessments 3x a year.

- b. Progress monitoring
  - i. Teachers just now seeing value of progress monitoring in forth year of implementation.
  - ii. Reading Coach's unsure how well others will embrace without teacher support and encouragement to engage in progress monitoring
  - iii. Teachers may be apprehensive about progress monitoring due to fear of results.
- ii. Reading Coach efforts to increase teachers' value or use of DIBELS
  - 1. Teachers feel validated when Reading Coach and other support personnel work with the teacher to better understand why a student did not perform well (e.g., understanding a student's home life, or conditions of the testing session, etc.).
  - 2. Teachers find value in data analysis when they see direct link to making plans for improving performances of students that involves others so that the teacher is no alone.
  - 3. Holding individual conferences with teachers and grade level teams have helped teachers find value in using the DIBELS through Reaching Coach support and encouragement.
- c. Teachers' Knowledge of DIBELS
  - i. Sharing research on correlation between DIBELS and FCAT helped increase value of DIBELS among teachers.
  - ii. Many teachers are still seeing the DIBELS as another high stakes test rather than a progress monitoring tool to guide instruction.
  - iii. Reading Coach perceptions of student reactions to being tested with DIBELS
    - 1. Reading Coaches validated teacher concerns that many students (especially KG students) do not test well with adults they are not familiar with.
      - a. Changes implemented to deal with this that have been found helpful:
        - i. RC's or assessors spend time in the classrooms more prior to testing cycles and engage in student activities.
        - ii. Provide students prior to testing what they will be asked to do and what to expect.
    - 2. Reading Coaches try to explain to KG teachers that the first cycle is not a reflection of their teaching, but rather more of a measurement of students' incoming knowledge and ability to follow directions.
    - 3. Reading Coaches feel teachers still feel pressure that they will be held accountable for the first DIBELS cycle with KG students despite Reading Coaches reassurance that it is not.

4. 1<sup>st</sup> grade students reactions to NWF – most get used to it quickly by the mid to end of the first grade and beyond – no longer an issue of trying to “make it a real word”.
  5. When training sheet for NWF subtest – called “SIM and LUT page”, most students usually say immediately before directions are given, “SIM/LUT” because they’ve seen it before.
  6. RCs concerned about the direction of NWF offering student option to say sounds or whole word because many first graders will often sound out and then also say whole word – this loses them time.
  7. Are they doing this because it is being modeled to them in the instructions?
- iv. Specific Subtests - NWF
1. RC’s validated teacher reactions to the NWF test as a test that should not be used.
  2. Teachers say to RCs that students are trying to make them real words by reading with long vowel sounds because they’ve been taught to use that strategy when trying to read/decode a word they’ve not seen before.
  3. RCs see a larger pattern where in the curriculum students are being introduced to final-e patterns and long vowel sounds mid-year. Most first graders are showing decreased performances during this mid cycle of DIBELS on NWF because of long vowel vs. short vowel.
  4. RCs try to educate teachers about NWF subtest by emphasizing value of decoding as well as sight word reading skills.
  5. Better that a student demonstrates a strategy that is consistently applied but wrong, instead of no strategy at all – just guessing.
  6. RCs have introduced “word work” activities that have helped teachers support student learning of decoding skills and strategies.
  7. Over the years, students become very familiar with the nonsense words and no longer becomes an issue for most of them.
  8. Questions about students in green or blue range of performance if they only give letter sounds instead of whole, blended, words – use raw data sheets to identify this. Value in the new addition by FCRR to include measure of whole words read on NWF subtest.
  9. The raw data for NWF is invaluable to see how the student attacks a word (e.g., segment all sounds and then blend whole word; segment initial sound and then read whole word; or just read whole word, etc.).
  10. Adding new measurement of number of words read still not as valuable as teachers looking at patterns in the raw data.
- d. Collecting and Using Assessment Data
- i. Data Collection Procedures

1. Teachers feel more comfortable and trusting of the DIBELS data when the same person collects the data each cycle for their class.
  2. Some described a process of either the whole class coming to the media center or DIBELS assessors pulling students one at a time to nearby quiet rooms for testing.
- ii. Reading Coach perceptions of teachers using DIBELS
1. 1<sup>st</sup> year vs. 4<sup>th</sup> year of implementation – evolution of teacher acceptance and use
  2. In beginning, teachers accepted DIBELS data that validated their expectations of a student and did not value DIBELS data for students who did not perform well on it.
  3. Took long time for teachers to see value of DIBELS (approx. 2-3 years).
  4. Some teachers find the DIBELS more reliable because someone else conducts the assessment – a sense it is more objective.
  5. In 4<sup>th</sup> year of implementation teachers trained to give some subtests of DIBELS to some students – many found more value in DIBELS through this. Doing it helps them understand it more.
- iii. DIBELS vs. other assessments
1. People administering the DIBELS need to be taking detailed notes about the children they've tested or some kind of indicator to know to go to the teacher and give them lots of information beyond the score. Going back to the teacher is one of the best ways to better understand why a student may not have performed well on the test. It's additional information (quantitative + qualitative).
  2. RCs all agree that sharing the raw DIBELS scoring sheets help make the data more valuable for teachers.
  3. Without looking at the raw data, the color itself may not indicate what to do to help that student.
  4. Using the raw data to find patterns is the role of the teacher – to use that information.
- iv. Teacher's Level of Proficiency in the use of DIBELS data
1. Teachers are inundated with so much data, but are unsure how to make use of it all at once.
  2. Variability exists in teacher willingness to take the next step in data analysis and data utilization – differences in reactions to change and new concepts.
  3. Teachers need much more support and training on how to use multiple sources of data to make decisions about a student's needs. (e.g., if student is low on DIBELS but is passing district assessment teacher will not provide interventions).
  4. Some teachers are still having difficulty seeing the correlation between fluency and comprehension.

5. Some teachers put too much emphasis on DIBELS at the expense of ignoring other measures/assessments and at the expense of only focusing on fluency.
- v. Views on the use of PMRN reports by teachers/administrators
1. Teachers needs more help to learn how to use the reports – more training.
  2. KG and 1<sup>st</sup> grade teachers often seem more proficient at using the reports.
  3. Some teachers are simply more proactive in using them and are more independent at accessing their own data online – teacher self-efficacy in using data.
  4. Relates to the evolution of Reading First grant: teachers are only now reaching a point where they are reading to begin embracing the use of the reports and using them to make instructional decisions.
  5. Data analysis is a less tangible process for many educators – need continued support for staff to reach a level of independence in this skill.
  6. RC's have to sometimes encourage teachers to use them by sitting with them one-on-one and consistently following up with them until they find the value in doing it themselves. This also relates the barrier of not enough time. Some teachers may have the skills but don't feel they have the time.
  7. Teachers will embrace using the reports when they see the value in them.
- e. Advice
- i. Concerns about the use of DIBELS
    1. Critical decisions being made on very small snapshots of student performance
      - a. It's just a snapshot and yet big decisions are being made on a snapshot (e.g., retention and special education consideration).
    2. In the Future
      - a. Concerns about teachers taking full ownership of conducting the DIBELS assessment out of fear that objectivity will be lost through bias or poor standardization.
        - i. Teachers in first grade have about 45 minutes to an hour that they could dedicate to testing each day (during reading center times). Even with that time, it would take more than a week to complete the whole class alone. Then add that time three for each type of test being done in the district – that is valuable instructional time for small group lessons that is lost.

- ii. Even with an assistant or sub, instructional time will be affected because the teacher still needs to plan for the sub/assistant ahead of time.
  - 3. Concerns about fidelity of use among non-Reading First schools.
- ii. Allow PMRN to generate graphs that reflect the correlation between students' oral reading fluency and later performance on 3<sup>rd</sup> grade FCAT.
- iii. Reading Coaches are needed!
  - 1. To make sure new teachers have the support they need to learn how to use the DIBELS effectively (i.e., turnover rates).
  - 2. To support teachers who are at different levels of understanding and proficiency in the use of DIBELS.
  - 3. Giving the DIBELS administration over to the teachers completely could threaten the validity as many teachers still see it as a test that they NEED to pass; or as a test that could be used against them regarding their effectiveness as a teacher.
  - 4. To coordinate data collection and analysis activities. No one else at the schools is currently trained to do this – a full time job.
- iv. If teachers are to take over using and coordinating the use of DIBELS, including the analysis and utilization of that data, they'll need:
  - 1. more planning time
  - 2. something needs to be taken away – district typically adds more stuff but rarely takes anything away. Teachers cannot be doing Common Assessments, Kaplan, DIBELS, and Project Focus assessments on their own. And then add more progress monitoring with DIBELS – it's too much. Something has to give.
  - 3. District needs to prioritize the assessments they are demanding teachers to use.
  - 4. Additional personnel in the classroom to sub or team teach to allow time for the teachers to give assessments reliably and analyze the data for use.
  - 5. Asking teachers to be responsible for testing students with DIBELS without taking other assessment requirements away is unrealistic and/or lead to valuable instructional time being lost.
- v. Need more progress monitoring and focus on the students growth in relation to interventions given.
- vi. Teachers and administrators need to pay more attention to a student's growth (i.e., score) in relation to standard rather than simply the color of the performance.

**V. Perceptions of DIBELS by “Specialists” (What attitudes and perceptions exist among persons other than teachers who participate in the collection, input, and analysis of DIBELS data throughout the school year?)**

- a. Culture/Climate of School
  - i. RF schools vs. Non RF schools
    1. Observed number of referrals for psychological evaluations have decreased as a function of using DIBELS data at RF schools.
    2. Observed that DIBELS is used to refer students for psychological assessments even when other assessment data indicates student is performing within average range at Non-RF schools.
    3. Day and Night differences in the amount of explicit and direct teaching of focused skills in KG and 1<sup>st</sup> grades (e.g., phonemic awareness) – diagnosticians are seeing big differences in students being assessed at RF and NRF schools on nationally norm-referenced achievement tests in reading.
  - ii. Climate for teachers right now very punitive and intense. Many teachers feel pressured to teach to the test out of fear of being judged professionally based on the DIBELS scores.
  - iii. Leadership plays a huge role in setting the right climate.
- b. Support/Resources Available
  - i. Role and Importance of Reading Coach:
    1. Provide training to DIBELS assessment team each year and in some cases before each cycle to ensure standardization procedures are known and followed for administering the DIBELS.
    2. Reading Coaches should have a stronger role in coordinating interventions with teachers for students.
    3. FCRR provided schools with large intervention binders with assorted collection of intervention ideas organized by skill area. Teachers don't have time to utilize this resource; RCs can help with this by being familiar with the activities and offering recommendations to teachers when they meet with them to go over the DIBELS data after each cycle.
  - ii. Teachers valuing the use of DIBELS data influenced by:
    1. Available time
    2. Access to color printer
    3. Motivation to use it
    4. Seeing the usefulness of it over time
    5. Teachers are overwhelmed by the amount of testing that is taking place
    6. Using graphs increases value for teachers, especially for progress monitoring data (line graphs). More "ah-ha" moments.
  - iii. Observed the following has led to more acceptance and increases in value of DIBELS by teachers:
    1. Follow up with teachers after assessment cycle to share results quickly.
    2. Share qualitative observations of students with the teacher as they relate to the DIBELS performance.

3. Show teacher actual DIBELS raw data/protocols
  4. Going into the classrooms before assessment cycles to give students advanced understanding of what they will be asked to do and given directions for how to participate in the assessment process.
- c. Teachers' Knowledge of DIBELS
- i. Observed that teachers at RF schools have evolved in their acceptance and use of DIBELS over time.
  - ii. Teachers concerned about it being a timed test and therefore not as valid a measure
  - iii. Concerns about the test being given by different people throughout the year who are unfamiliar with the students
  - iv. Reflections on the NWF subtest.
    1. Specialist observations do not necessarily agree or disagree with teacher comments on this subtest – most likely because of the unique background of the specialists (e.g., diagnosticians).
    2. Emphasized the importance of the Reading Coach to help teachers and staff understand the importance and usefulness of the NWF measure.
    3. At Non-RF schools, observed teachers who have backgrounds in reading have refused to use the NWF subtest or at least refuse using the results – find it has no value to read such words.
  - v. Sharing research on DIBELS/Reading Development
    1. Agree that this would help increase value of DIBEL among teachers
    2. It would provide more context for why the DIBELS is so useful.
    3. Administrators need to know this information just as well
    4. Sharing such information may not help them interpret the results better, but at least help them see the benefit of using it.
- d. Collecting and Using Assessment Data
- i. Observations that people collecting DIBELS data are more accurate when ongoing training and coaching are available – especially when assessment team members will be testing a different grade than before or it's been a long time since they worked with a specific grade level/subtest.
  - ii. Observations of assessing special populations (ESOL, Spec Ed, Speech, etc.) with DIBELS
    1. Helpful if same person keeps working with them who is familiar with them – to differentiate more accurately an error vs a foreign accent for example.
    2. Students who have special needs or circumstances may show a more true or valid performance with someone they know and have worked with before.

3. Sometimes it may be helpful to at least have a system in place that communicates specific student circumstances or characteristics that would need to be considered before working with such students regardless of who collects the information.
  4. Also recommended for progress monitoring since it is often not possible for the same person to do all the progress monitoring on a particular student because of schedules.
  5. Increases value of results when data and observations during testing are immediately shared with teacher.
- iii. Observations of procedures for collecting DIBELS at different schools:
1. Different at different schools.
  2. Procedures adopted by a given school not observed to change over the years, but become more efficient.
  3. Original choice of approach observed to be adopted by the influence of what seemed possible (i.e., staff available, schedules, structural arrangement of school).
- iv. Teachers conducting progress monitoring
1. 06-07 first year some teachers asked to give DIBELS to some students for progress monitoring.
  2. Observed teachers have problems with this because:
  3. Time available to engage in activity
  4. Teachers don't feel they get out of it as much as when someone else had done it and brings them the information and observation notes.
  5. Lack of suitable space to conduct the assessment without distraction or interruption – no one else to watch the other students
  6. Teachers at RF schools seem more ready to understand the role of progress monitoring than teachers at Non RF schools.
- v. Teachers administering DIBELS in absence of DIBELS assessment team or Reading Coach:
1. Unrealistic
  2. Teachers would need extra personnel regardless, to watch rest of class or else reliability and/or validity of results threatened.
  3. Similar problem to that faced by KG teachers currently with PIAP assessment
  4. Even if other kids are provided with independent activities, teacher conducting assessment alone with DIBELS could take weeks to complete whole class with allotted time available to teachers for such activities.
  5. Some have observed KG and 1<sup>st</sup> grade teachers who have taken initiative to do their own DIBELS assessments while intern covered class – these teachers did not require someone to come

- back to them after DIBELS cycles (when someone else is collecting the information) to help them interpret it.
6. Some teachers will be able to do this and be motivated to do this, but not likely that all teachers will – should be voluntary for teachers to do it rather than forcing people to do one more thing.
  7. Teachers in upper grades should be more encouraged to conduct their own DIBELS on at least those students who are not comprehending. Much less of the assessments in upper grades correspond with the content measured in DIBELS.
- vi. Teachers don't find value in DIBELS to drive instructional decision making.
    1. Teachers showing preference for Running Records even though not a fluency assessment – teachers need more information about the importance of fluency.
    2. Some teachers need to be more flexible about organizing student reading groups in the beginning weeks of school as the first cycle of DIBELS data is acquired. – this should be reviewed throughout the year as student's progress in their skills.
  - vii. Analyzing DIBELS data:
    1. Teachers need to put more emphasis on student growth rather than focusing only on the color.
    2. Educators need to be watching the student's trend rather than their single performance on one assessment cycle.
    3. Teachers needs more guidance and support/training on how to use multiple sources of data.

## **VI. Case Study Comparative Analysis Between Teachers and DIBELS Experts**

- a. Two Experts in the use of DIBELS/PMRN were separately and independently asked to review 1 KG case study and 1 First Grade case study.
- b. Each teacher asked to review case study matching the grade level they teach only.
- c. Both case studies reflected end of year cycle
- d. Each case study involved three specific PMRN reports chosen for use based on comments received by Reading Coaches and school staff who work with teachers and DIBELS.
  - i. Class Status Report
  - ii. Student Grade Summary Report (Box and Whiskers format)
  - iii. Reading Progress Monitoring Student Cumulative Report (referred to as Cumulative Report below).
- e. Each teacher/expert asked to give their impressions about the student case presented to them.
  - i. What would they do with a student like this?

- ii. What did they feel this student needs?
- iii. What other types of information would they want to know about this student?
- iv. How helpful are these types of reports?
- v. Are there other types of reports that are used or preferred?
- f. Expert reviews analyzed before teacher reviews
  - i. Results reported as aggregate data for experts and teachers, respectively by grade level.
  - ii. Results can be categorized into three main themes:
    - 1. Using the PMRN reports
    - 2. Additional information needed/wanted
    - 3. General comments about using DIBELS data
- iii. **Expert review of kindergarten case study**
  - 1. **Using PMRN Reports**
    - a. Using the cumulative report allows opportunity to see progress of skills
    - b. Using the cumulative report allows opportunity to see vocabulary level
    - c. Using the cumulative report allows opportunity to see scores on ISF in cycles 1-3
    - d. Using the cumulative report allows 1<sup>st</sup> grade teacher at beginning of year to see what students are coming in with – skills they have.
    - e. Using Class Status Report allows you to sort the data in different ways
    - f. Use reports to identify student strengths and weaknesses
      - i. Student is fluent in letter naming skills
      - ii. Student needs significant help in phonological awareness skills
      - iii. Student seems to have some letter-sound correspondence skills
    - g. Recommendations for the student in the case study
    - h. Using the Student Grade Summary Report allows for comparing student to class to standard
  - 2. **Additional Information Needed**
    - a. Need more information to understand why student is behind peers in ISF and PSF
    - b. Need more information
      - i. Is the student in ESE, ESOL?
      - ii. Does the student have any language impairments/delays
      - iii. Is English primary language
      - iv. How did student respond during testing
      - v. How does the student respond to activities/interventions in the classroom

- vi. What has been done for this student so far
  - vii. More background information
  - c. Need more information to understand why student went from 0 to 24 on NWF
  - d. Need more information to know what phonological skills the student does have - DIBELS data only tells us the student has a problem in phonological awareness skills.
- iv. Kindergarten teacher reviews of kindergarten case study**
- 1. Using PMRN Reports
    - a. All kindergarten teachers recognized and reported using the Class Status Report
    - b. All used the Class Status Report to identify student strengths or weaknesses
      - i. Student is “having trouble with sounds” based on the NWF
      - ii. Fluent in letter naming skills
      - iii. Student can do nonsense words but “doesn’t know sounds” referring to PSF
      - iv. “He knows his sight words”
      - v. Needs help with beginning sounds and letter-sound relationships
      - vi. Main concern by one teacher was PSF
      - vii. Recommendations for various activities to teach student
      - viii. Two teachers indicated this student should be referred for ESE consideration.
    - c. 3 of the kindergarten teachers reported or demonstrated using the Student Grade Summary Report to either identify student strengths or needs and/or use to explain to parents their child’s needs.
    - d. Only 1 teacher found value and used the Cumulative Report to identify student progression over the year.
  - 2. Additional Information Needed
    - a. Only 4 teachers asked for additional information
      - i. Two of the teachers wanted to know if student is in ESOL or has English as second language
      - ii. All 4 teachers wanted to know if student has any speech or language impairments
      - iii. 1 teacher asked for observations of student during testing (e.g., distractions?)
      - iv. 1 teacher asked if any short-term or long-term memory problems; other processing deficits
      - v. 1 teacher would want to eliminate questions about cognitive skills before assuming student isn’t trying.
- v. Expert review of first grade case study**
- 1. Using PMRN Reports

- a. Use Cumulative Report to see student progression through the year
  - b. Use Cumulative Report to see other outcome measures in vocabulary (PPVT) and comprehension (SAT-10)
  - c. Use reports to identify student strengths and weaknesses
    - i. Student needs help in phonics
    - ii. Student needs help in reading fluency
    - iii. Student strength in phonological awareness
    - iv. Student seems to be compensating somewhat for comprehension despite low vocabulary and low fluency skills
    - v. This case is an individual student problem because class is performing above benchmark higher than the student.
  - d. Don't just look at the colors, but also the numbers
2. Additional Information Needed
- a. What can this student do; what sounds do they know in phonics
  - b. Teachers need to use knowledge of reading development to figure out where to start helping a student (e.g., problem with fluency – is it because they can't decode words)
  - c. Teachers can do more progress monitoring
  - d. Teacher can access additional probes from Oregon website and give it themselves
  - e. FCRR website has several guides for teachers on the use of ongoing progress monitoring
- vi. **First Grade teacher reviews of first grade case study**
1. Using PMRN Reports
- a. All first grade teachers recognized and reported using the Class Status Report often.
  - b. All reported using the colors on the Class Status Report to identify student strengths and weaknesses
  - c. Only one of the 7 teachers valued using the Student Grade Summary Report and understood how to use it – but reported that the Box and Whiskers format was mostly used at the class level in her class.
  - d. All of the other 6 teachers reported no value in using Student Grade Summary Report – too confusing and too much visually.
  - e. A few teachers either recognized and/or found value in using the Cumulative Report – one teacher in particular used that report first to look at the student's progression over the year.
  - f. All teachers reported student strengths and needs:
    - 1. Student needs help in phonics
    - 2. Student needs help in reading fluency

3. Student strength in phonological awareness
4. Most teachers described various activities to teach the student phonics/oral fluency/sight word vocabulary
- g. The “Historical Report” and “Class Recommended Level of Instruction Report” were stated as alternative reports used by teachers.
2. Additional Information Needed
  - a. Only 5 of the first grade teachers inquired about additional information for helping to interpret the PMRN reports
    - i. Error patterns on the NWF?
    - ii. Is student and ESE student?
    - iii. Any disabilities (e.g., ADHD or Language Impairment)?
    - iv. Observations of student during testing (i.e., distracted)?
    - v. Is student ESOL or is English a second language?
    - vi. Error patterns on the ORF?
    - vii. What interventions have been tried already?
    - viii. What conditions exist at students home?
- vii. Expert Opinion on Using Data – general – comments about using data/DIBELS at Reading First schools.
  1. DIBELS data tells you there is a problem but doesn’t tell you how to fix it
  2. DIBELS is only one snapshot
  3. DIBELS Benchmark data not meant for instructional planning
  4. Having access to raw data/probes is valuable
  5. Progress monitoring data is more useful for making instructional decisions
  6. Frequency of progress monitoring depends on the student; generally more intensive cases need weekly; moderate risk could be bi-weekly or monthly.
  7. Only need to give one probe for progress monitoring
  8. Use data to determine if individual student problem or class-wide problem
  9. Teachers should use knowledge about student (e.g., background) to help them interpret DIBELS data
  10. Progress monitoring not happening nearly as much as would like to see
  11. Experts observe much variability among schools’ abilities to use data to guide instruction (not just with DIBELS).
  12. Agree that schools that have been doing it at least three years have a smoother process in place
  13. Experts observe that teachers need continued support for how to use data to make decisions about instruction – something not really taught in teacher training programs.

14. Reading Coaches are an invaluable source of support for professional development for teachers
15. Need to use DIBELS with other observations and other assessments
16. DIBELS data provides some information to differentiate instruction for class - broadly
17. Wouldn't base any decisions on just this DIBELS cycle
18. Making hypotheses about student deficits and using data to confirm or disconfirm