Philadelphia College of Osteopathic Medicine
DigitalCommons@PCOM

PCOM Psychology Dissertations | Student Dissertations, Theses and Papers

2019

Predictability of Curriculum-based Reading Measures for Statewide Test Performance

Rebecca Sarah Mark
Philadelphia College of Osteopathic Medicine

Follow this and additional works at: https://digitalcommons.pcom.edu/psychology_dissertations

Part of the School Psychology Commons

Recommended Citation

Mark, Rebecca Sarah, "Predictability of Curriculum-based Reading Measures for Statewide Test Performance" (2019). *PCOM Psychology Dissertations*. 480.
https://digitalcommons.pcom.edu/psychology_dissertations/480

This Dissertation is brought to you for free and open access by the Student Dissertations, Theses and Papers at DigitalCommons@PCOM. It has been accepted for inclusion in PCOM Psychology Dissertations by an authorized administrator of DigitalCommons@PCOM. For more information, please contact library@pcom.edu.

Philadelphia College of Osteopathic Medicine

Department of Psychology

PREDICTABILITY OF CURRICULUM-BASED READING MEASURES FOR STATEWIDE TEST PERFORMANCE

By Rebecca Sarah Mark

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Psychology

March 2019

PHILADELPHIA COLLEGE OF OSTEOPATHIC MEDICINE
DEPARTMENT OF PSYCHOLOGY

Dissertation Approval

This is to certify that the thesis presented to us by

_______________________________ on the ______ day of __________________,

20___, in partial fulfillment of the requirements for the degree of Doctor of Psychology,

has been examined and is acceptable in both scholarship and literary quality.

Committee Members' Signatures:

_______________________________, Chairperson

____

____

_______________________________, Chair, Department of Psychology

ACKNOWLEDGEMENTS

To my mom and dad, Joyce and Don, thank you for your constant support and love. You instilled in me the value of education and always encourage me to reach my full potential. You have always set an example for me of extraordinary determination and a relentless work ethic.

To my sister, Jessica, your incredible confidence and conscientiousness have always been inspiring for me. You are an amazing role model.

To Dr. George McCloskey, thank you for your patience and determination in guiding me through the dissertation process. I am impressed by your energy, humbleness, and extensive knowledge of reading and the psychology field. Your insight and compassion were appreciated.

To Dr. Virginia Salzer, your passion for research was instrumental in motivating my work on this dissertation. Your enthusiasm is contagious, and I have become enthralled by the experience of completing research, especially analyzing data.

To Dr. Lori Lennon, thank you for your intuitive feedback. Your thoughtful consideration of the data and your vast knowledge on the process of reading and on reading interventions had a monumental impact on this dissertation.

ABSTRACT

National legislation has led to an increasing need for school districts to demonstrate student reading progress using performance on statewide achievement tests as indicators of growth. This study added to previous research on the effectiveness of curriculum-based measurement (CBM) in predicting success on statewide reading achievement tests and determining whether a student is at-risk for poor performance on statewide tests. The current study analyzed the relationship between a CBM tool for assessing reading progress, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), and a statewide reading assessment, the Pennsylvania System of School Assessment (PSSA). This study compared the predictive efficiency of three components of the DIBELS, Oral Reading Fluency (ORF), Daze, and Total, for student performance on the PSSA. The study analyzed scores of 75 participants across and within Grade 4 and Grade 5. No significant differences were found between ORF, Daze, and Total scores or between fall and spring DIBELS administrations. Results indicate that ORF, Daze, and Total categories are similar predictors of student statewide test performance and that DIBELS Total category is not more predictive than individual DIBELS measures. Results also suggest that DIBELS is a valuable tool for school districts to monitor student reading progress.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Overview

Student performance on high-stakes statewide tests has far-reaching implications. Increased accountability for schools and early identification of students at risk for reading problems has become a national priority. Due to increased demands on schools to generate higher scores on statewide achievement tests, appropriate tools for monitoring student progress and identifying students who are at risk for failing statewide tests has become a concern (Hoffman, Jenkins, & Dunlap, 2009; U.S. Department of Education, 2001). Results of statewide tests often influence educational decisions concerning curriculum development and the allocation of resources (Shapiro, Solari, & Petscher, 2008). The use of curriculum-based measurement (CBM) tools could be an efficient means to monitor student progress, identify at-risk students, and provide appropriate interventions to increase student performance (McGlinchey & Hixson, 2004). Using CBM on an ongoing basis allows for the facilitation of program changes throughout the school year, rather than waiting until the end of the year when annual assessments are completed.

CBM is a form of assessment that uses standardized methods to determine student performance on aspects related to the curriculum (Ardoin & Christ, 2009; Hintze, Callahan, Matthews, Williams, & Tobin, 2002; Marcotte & Hintze, 2009). Student performance is considered among an established standard of performance and predetermined factors. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is a CBM tool used by school districts to assess reading performance while monitoring student levels of proficiency and risk for failing statewide tests (University of Oregon Center on Teaching and Learning, 2009). Students in Pennsylvania are administered the Pennsylvania System of School Assessment (PSSA) beginning in the spring of third grade. The use of CBM to track performance in preparation for statewide assessment is part of the Response to Intervention (RTI) model, referred to as Response to Instruction and Intervention in some states, employed by school districts (Fuchs & Fuchs, 2006, 2009; Mesmer & Mesmer, 2008).

RTI, a multi-tiered intervention model, provides responsive research-based curriculum, instruction, and assessment that may be used by school districts to improve student performance on statewide tests and identify at-risk students who may not achieve a Proficient level on the statewide tests (McGlinchey & Hixson, 2004; Shapiro et al., 2008). With the information and data generated by these interventions and measures, students can be provided with reading interventions at an earlier age in order to help them achieve success.

Statement of the Problem

Due to national legislation, there is increased pressure on schools to produce higher scores on statewide tests (Bursuck & Blanks, 2010). Statewide assessments provide information on student proficiency in academic content and skill areas and inform educational decisions. As a result of high-stakes assessments, there is a necessity to determine effective means of progress monitoring and the predictive efficiency of CBM for statewide test performance. It would be beneficial to obtain reliable indicators of reading achievement throughout a student’s progression of reading skill acquisition and to monitor how well at-risk students respond to interventions to enable efficient responses to any lack of progress or any skill deficit (Shapiro et al., 2008). There is a need to explore the use of appropriate CBM methods of monitoring student progress, identifying at-risk students, and providing effective interventions to increase student performance (McGlinchey & Hixson, 2004).

Many school districts throughout the United States have implemented benchmark assessments and progress monitoring in areas of reading, including oral reading fluency (ORF) and reading comprehension. The RTI model has been implemented in many school districts with success because this model helps to identify at-risk students and provide appropriate interventions (Fuchs & Fuchs, 2006, 2009; Mesmer & Mesmer, 2008). The RTI model is associated with improving reading proficiency, but research has yet to fully explore how reading CBM assessments can best be utilized within the model. More research is needed to analyze the predictability of CBM for statewide test performance. Determining the predictive efficiency of CBM contributes to a better understanding of how to improve early identification of at-risk students and implementation of appropriate interventions within an RTI framework.

National legislation has led to increased accountability of school districts for demonstrating student reading progress. Research has suggested that CBM is effective in predicting success on statewide reading achievement tests and determining whether a student is at risk for poor performance, which is beneficial in guiding early intervention development and strategies (Barger, 2003; Hintze & Silberglitt, 2005; McGlinchey and Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Wood, 2006). The current study compared the predictive efficiency of components of the DIBELS for elementary school student reading performance on the PSSA.

Purpose of the Study

The purpose of the current study was to examine the effectiveness of the DIBELS in predicting reading performance on statewide testing. The RTI model is linked to increasing levels of reading proficiency, but more research is necessary to support the use of the DIBELS within the RTI framework for advancing reading achievement. This study focused on the reading performance of an elementary student population in a school currently using an RTI model. The study analyzed the predictability of student DIBELS for reading performance on the PSSA. This study added to previous research concerning the utility of reading progress monitoring tools and has implications for guiding intervention.

CHAPTER 2

LITERATURE REVIEW

National Legislation

The passage of the No Child Left Behind Act (NCLB) in 2001 and the reauthorization of the Individuals with Disabilities in Education Improvement Act (IDEIA) in 2004 led to an increase in research devoted to reading instruction (U.S. Department of Education, 2001). NCLB is a campaign that is closely related to the statewide performance standards set for all public schools. This act was passed to ensure that all students are effectively involved in the learning process and achieve the established performance goals and standards (Bursuck & Blanks, 2010). Primary components of NCLB, the reauthorization of the Elementary and Secondary Education Act of 1965, include mandatory yearly testing of all students, state-established standards, increased accountability, empirically supported research for curriculum and interventions, an increase in funding flexibility, and parent choice for education (Hoffman et al., 2009; U.S. Department of Education, 2001). One goal of NCLB was to close the achievement gap between low performing and high performing students (U.S. Department of Education, 2001; Wiley & Deno, 2005).

NCLB was replaced with Every Student Succeeds Act (ESSA) in 2015 (U.S. Department of Education, 2015). ESSA was signed into law by President Obama in December 2015. Over time, the NCLB’s prescriptive requirements became increasingly unworkable for schools and educators. As a result of concern from educators and families, the Obama administration created a law that focused on the clear goal of fully preparing all students for success in college and careers. The ESSA includes provisions aimed at ensuring success for students and schools.

IDEIA and NCLB highlighted the need for the use of evidence-based instruction and interventions to monitor student progress. IDEIA implemented the RTI model, which allowed students to be identified as having a learning disability after failing to make progress in response to evidence-based instruction and interventions (U.S. Department of Education, 2004). This method of identification differed from the traditional model, which consisted of identifying a discrepancy between the student's ability and achievement levels. The employment of the RTI process as a result of IDEIA supported the concept of student performance improving through research-based interventions (Mesmer & Mesmer, 2008). This method also included collecting data that demonstrates the student has been provided appropriate instruction and repeated assessments of achievement.

As a result of IDEIA and NCLB, public schools are mandated to conduct yearly assessments to identify at-risk students, implement interventions, and track student progress (Schilling, Carlisle, Scott, & Zeng, 2007; Shapiro et al., 2008). Persistent efforts to improve reading performance among students in United States schools have continued over the last 60 years, even though 44% of fourth grade and 46% of eighth grade students have failed to meet the standards for reading proficiency, according to the Nation's Report Card in 2015 (Kendeou, McMaster, & Christ, 2016). The standards for reading performance include word reading fluency and the identification and use of meaning in both explicit and implicit forms. Students who experience reading difficulty are likely to struggle throughout their educational life and into employment as an adult.

If students do not make progress as measured by benchmark assessments and standardized testing measures, interventions must be implemented to remediate academic concerns, such as reading problems (Wiley & Deno, 2005; Wood, 2006). The frequency of monitoring early literacy skills has significantly increased since the passage of these two acts and become standard procedure in elementary school (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001; Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002; Stage & Jacobsen, 2001). This includes assessment of reading skills during the school year with benchmark measures and progress monitoring by means of frequent assessment, identifying students at risk for reading difficulties, and providing instruction and interventions to help students achieve proficiency levels (Buck & Torgesen, 2003; Crawford, Tindal, & Stiebar, 2001; McGlinchey & Hixson, 2004; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shaw & Shaw, 2002). The RTI framework supports evidence-based intervention and monitoring of reading progress.

Response to Intervention

Overview. The RTI model came about as a response to legislation and increased accountability for schools and as a solution to deficits in the process of identifying learning disabilities. RTI was first implemented in 2004 in response to national legislation for data collection and decision-making (Fuchs & Fuchs, 2006, 2009; Mesmer & Mesmer, 2008). The groundwork for the RTI model was laid in the 1970s, when schools were not effective in meeting the needs of students with reading disabilities. In 1976, the federal government started the initiative to improve the special education response for disabled learners.

Also in response to legislation, the government appointed committees to do research to identify and categorize children with learning disabilities (O'Connor, Swanson, & Geraghty, 2010). This research found that children with special learning needs comprised more than 17% of all students enrolled in public schools (O'Connor et al., 2010). Many of these students struggled with low grades and below average scores on state tests and often dropped out of school. Use of the RTI model has the potential to help students such as these to achieve a Proficient level of performance. The advancement of the RTI model highlighted the need to restructure the learning process for special education students in order to help them meet achievement standards (O'Connor & Klingner, 2010).

The National Center on Response to Intervention (2018) and the National Center for Learning Disabilities (NCLD) (2018) define RTI as using assessments and interventions to maximize student achievement, identify at-risk students, monitor progress, and provide research-based interventions. The RTI model is primarily used to identify students at risk for academic difficulties or behavior problems. Goals of the RTI model include providing early support to students, decreasing education costs, reducing inconsistencies in identifying special education students, preventing school failure, ruling out poor teaching strategies, and increasing the amount and quality of instruction. Once at-risk students are identified, interventions are implemented, and academic progress is monitored.

RTI is a systems-level tiered approach that provides targeted evidenced-based instruction and intervention for all students who may or may not be categorized as disabled (Fuchs & Fuchs, 2009; Hale, 2006; Tomlinson, 2011). The RTI model of programming is based on a problem-solving model that utilizes data to make educational decisions and requires a collaborative team approach from educational staff members. Three tiers of increasingly intensified instruction and intervention comprise the RTI system. The first of the three tiers is the general education curriculum, which aims to meet the needs of most students, or about 85% of the student population (Fuchs & Fuchs, 2009; Hale, 2006). All students are assessed with a universal screening tool at this level to determine if further individual attention to learning is necessary.

The secondary tier is for about 10% of the student population and involves more targeted instruction and interventions, typically administered in small group settings for students identified as at risk for learning difficulties (Fuchs & Fuchs, 2009; Hale, 2006). The tertiary tier addresses about 5% of the student population and targets specialized and individualized student needs. This tier includes students who did not respond to instruction and intervention at the secondary tier and often requires evaluation for special education services. Movement through the tiers is determined by progress monitoring, which consists of repeated assessments using CBM and comparisons to benchmark scores at each grade level.

Benefits. The RTI model is a positive means of restructuring the process of reading problem identification and instruction. Research shows that RTI interventions have had a positive impact on educational programs (Feifer, 2008; Fletcher et al., 2011). Schools that implemented RTI in 2011 were able to achieve more than 50% improvement in student performance on statewide assessments (Fletcher et al., 2011). RTI helps to ensure that students who are misidentified as having a learning disability will not be unintentionally forgotten (Burns, Scholin, Kosciolek, & Livingston, 2010). There is no uniform educational policy for teaching at-risk readers, which leads to accidental neglect of these students (Aaron, Joshi, Gooden & Bentum, 2008; Broxterman & Whalen, 2013). The RTI model creates a framework for educators to identify students who are at risk for poor performance in early grades and provide specialized instruction that closely aligns with developmental learning periods for reading (Russo, Tiegerman-Farber, & Radziewicz, 2009). RTI is beneficial in preventing students with reading concerns from being overlooked.

The RTI model helps educators to differentiate between a learning disability and an instructional disability. Research shows that 20% of students identified as having a reading learning disability do not actually meet learning disability criteria (Burns et al., 2010). A learning disability refers to hardwired deficits that can be modified in early developmental stages. An instructional disability refers to a lack of exposure or instruction (McCloskey, 2016). If certain reading skills are not taught, then the student's difficulties could be indicative of an instructional disability. Students wrongly labeled as having a reading disability may have actually not received effective instruction and may fall further behind (Wanzek & Vaughn, 2011). Using the RTI model reduces the risk of misclassifying a student as having a cognitive learning disability and differentiates between students who are doing poorly because of poor and inconsistent instruction and those who do poorly because of a cognitive learning disability (Powell, Higgins, Aram, & Freed, 2009). RTI helps educators to discern the complexities between abilities and possible lack of instruction in certain reading areas.

The tier system allows for data-based decision making. With a standard method of identifying students with concerns, students are usually not helped until they are starting to struggle, whereas an RTI approach uses universal screening to indicate a problem area early on. The model provides specialized instruction to all students who need it. When students are not responsive to general instruction, the model provides a means for considering a learning disability as a factor responsible for the discrepancy and lack of response, which may then indicate the need for special education support (Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010). The RTI approach reduces the amount of time that students have to wait to receive necessary instruction and intervention. Educators can base their decisions and strategies on data rather than assumptions.

The positive effects of intervention facilitated by using an RTI model can be examined over time to observe the rate of student improvement. Wanzek and Vaughn (2007) found that Tier 3 interventions provided for 100 or more sessions, equivalent to 20 weeks of daily intervention, were associated with positive student outcomes. More recently, Wanzek and Vaughn (2011) examined the implementation of an RTI model over a 5-year span. Results indicated a trend of fewer students identified for special education with each year of implementation. Teachers were provided with professional development in reading, which may have contributed to the study's outcomes (Wanzek & Vaughn, 2011). Tran, Sanchez, Arellano, and Swanson (2011) conducted a meta-analysis and found no significant moderating effects of duration of intervention and number of sessions on measures of reading. The average number of weeks of intervention was 18.86, and the average number of sessions for intervention was 31.68 (Tran et al., 2011).

Al Otaiba et al. (2014) found positive effects of an RTI model obtained over 1 year. A typical RTI model utilizing decision rules that waited to assess response to Tier 1 was compared to a dynamic RTI model in which Tier 2 and Tier 3 interventions were implemented immediately, when necessary. Dynamic RTI allowed for student movement across tiers every 8 weeks and for students to receive Tier 3 interventions when necessary, rather than requiring them to first go through Tier 1 and Tier 2. Students in the dynamic RTI had higher reading performance at the end of the year span, and the positive effects accumulated across the year (Al Otaiba et al., 2014).

Simmons et al. (2015) also found positive outcomes of an RTI model over 1 year. Tier 2 interventions implemented in kindergarten were examined. Positive effects were found for students whose intervention progression was adjusted every 4 weeks, based on data on student mastery of skills (Simmons et al., 2015).

Wanzek et al. (2016) examined Tier 2 interventions implemented in Grades K-3 through a meta-analysis and found positive outcomes of Tier 2 interventions on foundational reading skills, with no differences in effects related to the number of intervention hours. Tier 2 interventions are typically provided for 15 to 99 sessions over approximately 4 to 32 weeks. Results indicated that sessions were implemented for a range of 4 to 80 hours, with sessions of 30 minutes occurring most frequently (Wanzek et al., 2016).

When considering the time frame of RTI, adjustment and flexibility are key factors (Al Otaiba et al., 2014; Simmons et al., 2015). Duration of intervention and sessions vary within and between tiers, and there appear to be no moderating effects of duration of intervention and number of sessions on reading outcomes (Tran et al., 2011; Wanzek & Vaughn, 2007; Wanzek et al., 2016). Positive effects of an RTI model on student reading outcomes are observed over a school year (Al Otaiba et al., 2014; Simmons et al., 2015; Wanzek & Vaughn, 2011). Adjustments every 4 or 8 weeks as well as fluid movement through tiers are beneficial to student success (Al Otaiba et al., 2014; Simmons et al., 2015).

The benefits of RTI can impact student performance over time with individualized and adaptive support. Poor performance at lower grades has lasting effects and continues to be a concern through high school (Woolley, 2011). Research shows that an RTI model has the capacity to help educators improve learning for all students by preparing them for college and the critical thinking that is required at all stages of life (Aaron et al., 2008; Burns et al., 2010). The RTI model helps educators identify reading problems early and eliminate these problems in the initial stages in a student’s education during the early grades, which improves student success in general and on statewide achievement tests (Berninger, Abbott, Vermeulen, & Fulton, 2006; Block, Parris, Reed, Whiteley, & Cleveland, 2009). The intensity of RTI is adjusted according to student performance, improvement, and responsiveness and provides an opportunity for every student to improve (Wooley, 2011). Students who are performing poorly are identified and given appropriate interventions to help them improve at an early stage, which has positive effects for the subsequent years of the student’s education.

RTI supports identification and intervention tools that are helpful in meeting legislative demands. With the NCLB Act, the Department of Education has mandated rules and regulations concerning student progress on statewide achievement tests and educator accountability for student progress (Burns et al., 2010; Tomlinson, 2011). The Department of Education has also advocated for the use of CBM, which sets standard performance in statewide tests and acts as an assessment for identifying student learning status according to the learning and performance standards (Elbaum, Arguelles, Campbell, & Saleh, 2004). In addition, a stipulation of NCLB was that states must test all public school students in reading and math every year from third through eighth grade and then again in high school, which results in a loss of government funding if state standards are not achieved (Whitten, Esteves & Woodrow, 2009; Woolley, 2011; VanDerHeyden & Burns, 2010). As a result of these policies, it is important for schools to consider the implementation of RTI to identify students who are at risk for failing statewide tests.

Implications of implementation. The implementation of RTI relies on many factors that influence the efficiency of the model. Educator response to RTI and CBM is one factor influencing the efficacy of the system. Research has shown that educators report favorable results using RTI. Educators believe that RTI interventions are more efficient than CBM and that the two methods used together improve the learning experience for all students (Feifer, 2008). The greatest benefit of RTI is that this intervention model helps students feel better about their success and achievement with regard to exams (Roehrig et al., 2008). In addition, parents are usually satisfied in general with RTI because it provides appropriate help to their children promptly.

The RTI model provides feedback to all stakeholders, including parents, regarding the instructional and learning strategies that teachers are applying to help children (Roehrig et al., 2008). Teachers are usually satisfied with the RTI model because this intervention allows them to more efficiently and easily identify problem areas and help students achieve to the best of their ability within the general education system. Teachers are able to adopt and implement the most efficient interventions in order to address a student's problem, and other school staff members may also assist in helping the student and improve overall school performance.

Another factor affecting implementation is accurate and effective instructional strategies. The RTI model identifies precise and detailed instructional strategies that may specifically benefit a particular student and provides useful information necessary to develop subsequent interventions (Roberts, Good & Corcoran, 2005). This model also provides information regarding how teachers may restructure their instructional strategies in order to fit specific student learning requirements for students who are at risk of failing and to intervene early, taking the initiative to help struggling learners at the earliest stages of the reading skill acquisition process (Roberts et al., 2005). The RTI model places an emphasis on treatment validity and follows through on the remediation process by suggesting the best instructional and teaching strategies that may be adopted by teachers (Peterson & Shinn, 2002).

The RTI model provides an opportunity for teachers to continually monitor the performance of students using a multi-tier reading approach throughout their academic careers and to restructure teaching strategies according to improvement and progress. All stakeholders get involved in learning and progress monitoring (O'Connor & Klingner, 2010). Teachers, the school principal, parents, special education specialists, tutors, the reading specialist, the school psychologist, social workers, and the student all share the responsibility to help every child to achieve and meet their full academic potential in school by collaborating, analyzing, designing, planning, and implementing efficient instructional strategies (Pullen et al., 2010).

Another aspect that influences implementation is the acceptance of the model. This model must be fully accepted and supported by all the stakeholders in order for it to be successful in a school district (Mellard, McKnight & Woods, 2009; Pullen et al., 2010). Gaining overall support for the model may be challenging, and a lack of support may limit the effectiveness of the RTI model because it requires teamwork for implementation (Nelson, 2008). An RTI team has the responsibility to identify the problems and issues and select the most efficient solutions, and to identify the opportunities and challenges, based on school needs, resources, and specific student deficits (Pullen et al., 2010). A support team familiar with the implementation of an RTI model outside of the school district may also be necessary to help the school team and maximize the potential for the successful implementation of the RTI model (Nelson, 2008).

Research shows that the RTI model has been effective, based on the positive response of students and improvement in their performance following the implementation of the RTI model across multiple tiers (Nation & Angell, 2006). Through collaboration, the RTI model allows teachers to use research-based and evidence-based instructional strategies throughout a multi-tiered service delivery system to improve performance for students who have disabilities, as well as for students who require different types of instructional strategies (Ehri, 2005; Langdon, 2004; Nation & Angell, 2006). In addition to successful implementation of the RTI model as a means of addressing the demands of national legislation regarding statewide testing, the use of CBM is helpful in monitoring student performance.

Curriculum-based Measurement

The use of CBM is an important element of the RTI model and provides a way to monitor student performance in consideration of statewide testing performance. CBM was developed in the 1970s and 1980s as an alternative to standardized tests and informal teacher observation (Deno, 1985). Standardized tests do not always align with curriculum objectives, and the reliability and validity of informal observation by teachers has not been established. The objective of CBM is to monitor, evaluate, and modify instruction and measure student progress with brief measures of basic skills and standardized methods to assess fluency (Ardoin & Christ, 2009; Hintze et al., 2002; Marcotte & Hintze, 2009).

CBM is a form of assessment that uses standardized methods to determine student performance on aspects related to the curriculum. Student performance is considered in the context of an established standard of performance and predetermined factors. CBM is considered effective for monitoring academic skills over time in a systematic manner (Klingbeil, Van Norman, & Nelson, 2017). CBM is a useful tool in providing information on instructional decision-making by assessing the performance of a student within the curriculum and determining the effectiveness of current instructional methods (Sattler, 2014).

There are many advantages of using CBM in a school district. Using CBM on an ongoing basis allows for the facilitation of program changes throughout the school year, rather than waiting until the end of the year when annual assessments are completed. Using a reading CBM to screen at-risk students may be a better measure of total reading achievement than group-administered norm-referenced achievement tests, which are more costly and time consuming (Ardoin et al., 2004). The use of CBM to predict performance on statewide testing may also be beneficial in identifying students at risk of failing the tests and in guiding interventions (Barger, 2003; Shaw & Shaw, 2002; Stage & Jacobsen, 2001).

Effectiveness of CBM is influenced by the number of administrations of CBM and teacher interpretation of results. Educators administer alternate forms of CBM over time and examine trends in order to observe patterns. When interpreting CBM results, educators can examine changes and consider if the variations are attributable to changes in reading rate or measurement error (Klingbeil et al., 2017; Van Norman & Christ, 2016). The accuracy of predicting future performance based on data from CBM measures is dependent upon the number of data points. The number of data points depends on the frequency of administration of the CBM measure. Research has found that decisions based on numerous data points, such as 20 data points, or over a long duration, such as over 12 to 14 weeks, improve accuracy in prediction (Klingbeil et al., 2017; Van Norman & Christ, 2016). The duration of data collection is the most prominent factor affecting accuracy in prediction. Decisions concerning resulting interventions are based on visual analysis of data or comparison of a student’s results to an expected goal.

Research has found that teachers are proficient in comprehending data from CBM measures, but may struggle with interpreting and linking data to instruction. Teachers often have to communicate their perceptions about student academic performance, and these judgments are often based on informal observation (Feinberg & Shapiro, 2003). It is possible that a teacher’s impression of a student’s academic performance may then influence expectations and interactions with the student. It is important that teachers receive appropriate training in this area (van den Bosch, Espin, Chung, & Saab, 2017). It would be helpful for teachers to be given results, along with consultation regarding suggested instructional changes (Graney & Shinn, 2005).

It is important that CBM data be interpreted accurately (Klingbeil et al., 2017; Van Norman & Christ, 2016). Incorrect decisions as a result of misinterpreting CBM reading data may lead to the continuation of ineffective instructional practices or interventions. Visual analysis supplemented with trend and goal lines is effective, but is still prone to incorrect interpretations. Additional data is helpful in making more accurate interpretations of CBM reading results. Also, following the directions provided for administration and scoring the CBM measure and collecting data in settings free from distractions helps reduce variability in observation.

Specifically with regard to reading, CBM scores are considered outcome measures of reading competence (Klingbeil et al., 2017). Research supports the reliability and validity of CBM for reading (Hosp & Fuchs, 2005; Klingbeil et al., 2017). These measures may be used to screen, identify, and monitor reading problems with specific skills including ORF, decoding, word reading, and reading comprehension throughout the grades (Hosp & Fuchs, 2005). Best practices for using CBM reading tools most effectively have yet to be determined. Although there are numerous CBM tools designed to measure reading ability, student performance varies within grades and across tools (Ford, Missall, Hosp, & Kuhle, 2017). Educators should have consistent direction for measurement selection and use.

Reading

CBM can be used to assess components of the complex process of reading. The major components of a Balanced Literacy model of reading are phonological awareness, decoding, fluency, vocabulary, and comprehension (Lennon, 2017; McCloskey, 2016; National Institute of Child Health and Human Development [NICHHD], 2000). Balanced Literacy refers to instructional methods and programs that are aimed at addressing most of the elements that comprise a brain-based cognitive neuropsychological model of reading (McCloskey, 2016). A neuropsychological perspective focuses on a multifaceted model of cognition and learning and considers a student's reading performance by exploring the relationship between reading abilities, processes, skills, and lexicons, rather than categorization of cognitive abilities and academic skills (McCloskey & Perkins, 2013; McCloskey, Perkins, & Van Divner, 2009). A comprehensive neuropsychological assessment explores the relationship between the brain and behavior. A neuropsychological perspective emphasizes the progression of reading and highlights the necessity to focus on reading skills and how to teach these skills.

Metacognition and executive functions also play a role in the reading process (McCloskey, 2016; Zabrucky, Moore, Agler, & Cummings, 2015). Metacognitive knowledge includes knowledge of cognitive abilities, tasks, and strategies and develops in response to a variety of experiences and instruction. Metacognitive experiences involve the active processes of assessment, such as checking one's understanding, as well as strategy use, such as testing oneself, and overlap with the concept of self-regulation.

Experiences and self-regulation involve reflection and action and are important to learning (McCloskey, 2016: Zabrucky et al., 2015).

Executive functions cue, direct, and coordinate the mental capacities necessary for reading (McCloskey, 2016; McCloskey & Perkins, 2013). Executive functioning skills, such as working memory, planning, organization, and self-monitoring, affect student ability to read (Cutting, Materek, Cole, Levine, & Mahone, 2009). Executive functions cue, coordinate, and direct attention for accurate perception and discrimination between letters and words, processes for word pronunciation, and production, prosody, and rate for efficient word reading. Prosody refers to rhythm and pitch when reading. These functions also involve coordinating the retrieval of information and use of attention and memory resources for reading words and connected text. Cueing retrieval of knowledge, the use of working memory resources, and the use of oral expression contribute to creating meaning for comprehension of text. Coordination of the use of strategies for reading words and deriving meaning from text is also necessary. Assessment and intervention of reading components should consider executive functions. Reading problems can result from or be worsened by ineffective or inconsistent use of executive functions (McCloskey & Perkins, 2013; McCloskey et al., 2009).

Reading is a multifaceted process involving the simultaneous use of multiple processes and typically requires the progression of acquisition of skills (Lennon, 2017; McCloskey, 2016; NICHHD, 2000). Students typically acquire phonemic awareness in preschool to second grade. Phonemic awareness is the ability to recognize and manipulate phonemes in spoken words (NICHHD, 200). Phonemes are the smallest units of spoken language. Phonemic awareness emerges in the preschool years and plays a causal role in learning to decode words (Catts, Herrera, Nielsen, & Bridges, 2015). Decoding refers to applying knowledge of letter-sound relationships to pronounce written words. Phonemic awareness is generally followed by decoding and orthography around first to second grade. Orthographic processing refers to the generation of internal visual representations of letters, words, and numbers. Fluent reading, which is reading text quickly and accurately, typically occurs around second to third grade (Meisinger, Bloom, & Hynd, 2010). Students begin to learn content area information requiring higher level skills, such as making inferences, drawing conclusions, and evaluating what is read, beginning around fourth grade (Kendeou et al., 2016). ORF is generally followed by the integration of vocabulary and comprehension skills.

Oral reading fluency.

Overview. ORF, defined as the number of words read correctly in 1 minute from connected text, involves effortless, smooth, and accurate reading. ORF is the ability to read a text accurately, quickly, and with expression (Armbruster, Lehr, & Osborn, 2001; Daly, Chafouleas, & Skinner, 2005; NICHHD, 2000; Wood, 2006). Difficulty with ORF skills may affect success in reading content areas. Strong ORF skills are necessary for comprehension of content area later in reading experience.

Components of ORF include sight word recognition, automaticity of recognition, reading speed, reading accuracy, and prosody (Lipka, 2017; McCloskey, 2016; NICHHD, 2000; Sabatini, 2002). Predictors of ORF include rapid automatized naming, phonological awareness, and working memory. Rapid automatized naming refers to quickly naming items and contributes to efficient automatic sight word recognition and fluent reading (Lipka, 2017; McCloskey, 2016). Phonological awareness is the ability to recognize and manipulate units of language. This highlights the relationships between letters and sounds. Working memory, the ability to process and store information, allows for decoding words quickly and retaining working memory resources for other reading components (Lipka, 2017; McCloskey, 2016). Working memory is involved in processing information while learning and producing.

Successful ORF is dependent upon various skills and processes at different stages of reading acquisition. Lipka (2017) examined underlying processes and predictors of ORF from Grades 2 to 6. Results indicated that phonological decoding was the most significant predictor of ORF at Grades 2 to 6. Rapid automatized naming contributed to ORF in Grades 2 to 5. Working memory was a predictor of ORF in Grades 2, 3, and 6, but not in Grades 4 or 5 (Lipka, 2017). Different factors play a more significant role in reading in different grades. Student supports and intervention strategies to help students acquire ORF skills are necessary (Miller & Schwanenflugel, 2008).

Measures. Tasks that measure ORF may be used to identify at-risk students and to monitor student progress. Some of these measures include running records, miscue analyses, informal reading inventories, qualitative reading inventories, and leveled reading passages (Armbruster et al., 2001; Hudson, Lane, & Pullen, 2005; Pikulski & Chard, 2005). Running records are means of documenting a student's individual reading of a text. Teachers may use some type of symbol to denote information while listening to a student read. Rapid automatic naming tasks can be used as baseline measures of ORF, as can sight word recognition tasks, and word decoding tasks may be used to examine sight word ORF and nonsense word decoding fluency (McCloskey & Perkins, 2013).

It is recommended that educators use 1-minute ORF measures to obtain information on student ORF (Fuchs et al., 2001; NCLD, 2018; NICHHD, 2000; U.S. Department of Education, 2001; Shapiro et al., 2008). These measures combine phonological segmenting, decoding, and rapid word recognition and assess accuracy and rate, along with sensitivity to growth over a short time. In a 1-minute ORF measure, the correct number of words read in 1 minute from a passage is calculated, which yields the ORF rate. The median of the ORF rates obtained from the reading of three passages three to four times a year becomes a student's benchmark ORF rate (Hintze et al., 2002; Shapiro et al., 2008). ORF measures include alternate forms of comparable difficulty.

CBM ORF tools have predictive utility and successfully differentiate between students with and without reading problems (Kilgus, Methe, Maggin, & Tomasula, 2014). CBM ORF tools' sensitivity ranges between .80 and .83, and specificity ranges between .71 and .73 (Kilgus et al., 2014). Variables that affect the efficiency of CBM ORF include the score used to define students considered to be at risk for reading problems, the time between the CBM and criterion test administration, and the percentile rank corresponding to the criterion test cut score. It is also important to consider examiner variation in administration and scoring (Cummings, Biancarosa, Schaper, & Reed, 2014). Educators administering CBM should receive in-depth training.

A type of CBM tool that assesses ORF is DIBELS. DIBELS is a science-based, outcome-driven model with formative assessments used to identify at-risk students and provide interventions and progress monitoring (Hoffman et al., 2009; Schiling et al., 2007). DIBELS is a CBM reading tool consisting of measures of early literacy skills, including phonological awareness, alphabetic principle, phonics, and comprehension, in addition to ORF (University of Oregon Center on Teaching and Learning, 2018). DIBELS Total score, also known as composite score, is a combination of the DIBELS measures, provides an overall estimate of literacy skills and/or reading proficiency, and will be referred to as DIBELS Total score in this study (Dynamic Measurement Group, 2010). DIBELS ORF (DORF) measures are intended to obtain benchmark and progress monitoring assessments that are equivalent to each other (Good & Kaminski, 2002). Increases in student scores should represent increases in student skills.

DORF is a standardized, individually administered test of accuracy and fluency with connected text (Good & Kaminski, 2002.) It is a standardized set of passages and administration procedures intended to assist in identifying students who may need additional instructional support. DORF administration requires the student to read three grade-level calibrated passages aloud for 1 minute each. The student's score, or ORF rate, is the median number of correct words per minute read aloud (Good & Kaminski, 2002).

DORF assessments, developed from research conducted at the University of Oregon's Center on Teaching and Learning (2009), predict literacy achievement and assist in the determination of the necessity of early intervention. The areas assessed with DORF are oral reading speed and accuracy (Riedel & Samuels, 2007). DORF assessments are empirically supported, predict literacy achievement, address oral reading speed and accuracy, and provide early intervention strategies (Riedel & Samuels, 2007).

Positive aspects of using DORF include efficient screening and progress monitoring and support for the three-tiered RTI model. In addition, DORF has a relatively short administration time and provides quick results that can be easily communicated to students' parents. Research shows that the DORF is widely used and is a valid and reliable measure of ORF (Goffreda, Diperna, & Pedersen, 2009; Hoffman et al., 2009; Rouse & Fantuzzo, 2006; University of Oregon Center on Teaching and Learning, 2009).

There are challenges associated with administration and interpretation of DORF. Research has found that DORF may overemphasize reading speed and requires concentrated time for individual administrations (Hoffman et al., 2009; Riedel & Samuels, 2007). Also, DORF provides quick results and does not assess reading comprehension, another important area of reading. This may lend to inaccurate information and an incomplete perspective of a student's reading performance, especially when DORF results are considered in making important educational decisions.

Interventions. Student difficulties with ORF can lead to trouble with learning and achievement in academic areas; therefore, the use of appropriate interventions is important to consider. Practice is an important element of increasing student ORF (Begeny & Martens, 2006). Students should read interesting texts that are at their grade level, read frequently, and reread texts. It is also helpful if parents are involved in helping students practice reading at home. Parent tutoring, along with reinforcement, student choice of intervention, and performance feedback, are other motivating factors associated with success of interventions (Daly & Kupzyk, 2012). A taped reading programming that includes listening passage preview, repeated reading, and performance feedback can be incorporated into parent tutoring at home (Kupzyk, McCurdy, Hofstadter, & Berger, 2011). Motivation, value of reading, and parental involvement influence student reading performance (Begeny & Martens, 2006; Siah & Kwok, 2010).

If parents are motivating and encourage reading at home, it is likely that students will value and practice reading.

An intervention that can be used to improve ORF is sustained silent reading (Siah & Kwok, 2010). Sustained silent reading is a period of uninterrupted silent reading in which students choose the books that they would like to read. Schools typically reserve approximately 15 minutes of each school day for sustained silent reading time. Sustained silent reading was developed in the 1960s and implemented in many public schools by the 1970s. Sustained silent reading is more effective when students place a high value on reading. Various sustained silent reading programs have been developed. These include Free Voluntary Reading, Drop Everything and Read, Daily Independent Reading Time, and Uninterrupted Sustained Silent Reading (Gardiner, 2001). All of these programs allow for uninterrupted silent reading, but may vary according to factors such as student interests. Sustained silent reading is aimed at fostering student enjoyment of reading, as well as improving ORF and other areas of reading.

Another intervention that educators can use to address ORF is repeated reading. Repeated reading practice is a reading fluency intervention program that involves rereading the same text in order to increase the rate and accuracy of oral reading (Algozzine, Marr, Kavel, & Dugan, 2009; Ardoin, Eckert, & Cole, 2008; Ardoin, McCall, & Klubnik, 2007; Musti-Rao, Hawkins, & Barkley, 2009; Rasinski, Homan, & Biggs, 2009). Students read a short passage several times. When utilizing repeated reading practice, a fluency criterion is set that may include correct words per minute and a specific number of errors (Yurick, Robinson, Carledge, Lo, & Evans, 2006). The student reads and rereads a passage until the fluency criterion is achieved, then begins this process with a new passage. Repeated reading can incorporate isolated word reading practice, unison reading, error correction, performance cueing, and feedback (Lo, Cooke, & Starling, 2011). This intervention has been found to be successful for students with and without identified disabilities (Begeny et al., 2010).

Readers' Theatre is a version of repeated reading that provides students who have varying reading levels of ability with scripts and specific parts that the students rehearse. The students then perform these scripts for other students. This intervention is a supplemental intervention to classroom instruction that is flexible, practical, engaging, and promotes reading confidence (Algozzine et al., 2009; Musti-Rao et al., 2009; Yurick et al., 2006).

Peer-mediated repeated reading has also been found to improve student ORF (Yurick et al., 2006). With peer-mediated repeated reading, two students take turns reading a passage to each other. They keep reading for a predetermined number of times or until they achieve a set fluency criterion.

Passage Previewing is another ORF intervention. Through the use of Passage Previewing, students read or listen to a passage (Begeny & Martens, 2006). They are then instructed or tested on that passage. One type of Passage Previewing involves the student listening to a more skilled reader reading a passage while the student follows along silently. Passage Previewing has been found to be effective for various populations of students.

Other means of improving ORF include rapid word naming, phase-drills, and fluency training in phoneme blending. Students can practice rapidly naming words and pseudowords (Begeny & Martens, 2006: Carnine, Silbert, Kame'enui, & Tarver, 2004).

With phase-drills, students read text from a passage repeatedly, with the addition of reading a specific phrase that contains a word that the student previously read incorrectly. Building fluency in phonemic awareness is also helpful because phonemic awareness is an important skill in ORF (Martens, Werder, Hier, & Koenig, 2013). Martens et al. (2013) found that training students to fluently blend phonemes while reading trained and untrained words in lists and passages improved ORF.

Interventions that address ORF, as well as other components of the Balanced Literacy model that aims to incorporate these elements, include Phonics for Reading, Six-Minute Solution, and Sonday System. These programs incorporate factors such as repeated reading, motivation, and peer monitoring. Phonics for Reading is an intervention that emphasizes ORF, phonemic awareness, and decoding through systematic teacher-directed lessons (Curriculum Associates, 2017). Six-Minute Solution is another ORF program that emphasizes phonics, sight word vocabulary, and repeated reading (Voyager Sopris Learning, 2018). Students work in pairs, small groups, or individually. Sonday System is a multisensory structured phonics, reading, writing, and spelling program that addresses ORF, phonological and phonemic awareness, consonants and vowels, spelling, vocabulary, and other elements (Winsor Learning, 2017).

Overall, repeated reading aloud, parental involvement, practice, and valuing reading are key factors in successful interventions (Begeny & Martens, 2006; Yurick et al., 2006). Consideration of the Balanced Literacy model of reading and executive functions is important to intervention success (Lennon, 2017; McCloskey, 2016). It is beneficial for students to be aware of their ORF abilities and understand how ORF fits into the process of reading. Intervention efficiency depends on individual student factors, instruction in strategy use, and flexibility in implementation. A combination of interventions is most effective (Begeny & Martens, 2006; Siah & Kwok, 2010; Yurick et al., 2006).

Reading Comprehension.

Overview. Reading comprehension is another important component of reading and the Balanced Literacy model (Lennon, 2017; McCloskey, 2016). Reading comprehension refers to deriving meaning from text through the combination of input from basic skills with stored knowledge and the use of language and other abilities to understand what is read (Kendeou et al., 2016; Lennon, 2017; McCloskey, 2016). Meaning of text is constructed through the interaction between the reader's knowledge and experience and the content of the text. Reading comprehension includes the visual process related to word reading, identity of phonological, orthographic, and semantic representations, and connecting words using rules of syntax to understand the underlying meaning of a sentence (Kendeou et al., 2016). To comprehend reading, an individual must integrate meaning across sentences, make use of relevant background knowledge, generate inferences, identify text structure, and consider an author's goals and motives. Individuals are required to read expository texts during formal education or while reading newspapers, magazines, and legal and medical documents in real world settings.

Reading comprehension is a complex cognitive process that requires higher level thinking skills and practice (Cutting et al., 2009; Yovanoff, Duesbery, Alonzo, & Tindal, 2005; Zabrucky et al., 2015). This process requires direct instruction, active engagement, and word recognition (NICHHD, 2000). Vocabulary, understanding explicit information, inferential comprehension, evaluative and critical reading, and emotional sensitivity all contribute to the comprehension of written text (Cutting et al., 2009; Yovanoff et al., 2005). Reading comprehension is considered a hierarchical framework that includes a continuum of basic to higher level skills that develop simultaneously and independently (Tarchi, 2015).

The integration of various cognitive skills, reading skills, and content knowledge skills is required for successful reading comprehension (Bashir & Hook, 2009; Cutting et al., 2009; NICHHD, 2000; Swanson & O'Connor, 2009; University of Oregon Center on Teaching and Learning, 2009). Cognitive skills necessary include active thinking, working memory, receptive and expressive language, reasoning with language, and visuospatial translation, or visual perception of the spatial relationships of language (McCloskey, 2016). Successful reading comprehension is accomplished by efficient use of mental resources. Reading comprehension is most effectively taught through the use of cognitive strategies.

Phonemic awareness, decoding, sight word recognition, fluency, accuracy, and vocabulary are components of reading necessary for comprehension (Bashir & Hook, 2009; Berninger et al., 2006; Cutting et al., 2009; Daly et al., 2005; Hosp & MacConnell, 2008; Russo et al., 2009). Content knowledge skills include syntax, sentence and paragraph relationships, and predictions. These skills are used simultaneously in order to comprehend efficiently. The performance and acquisition of reading comprehension skills are influenced by a number of variables including verbal cognitive ability, background knowledge, and instantiation of work knowledge, text structure, and efficient cognitive processes (Russo et al., 2009).

ORF is related to reading comprehension (Bashir & Hook, 2009; Berninger et al., 2006; Cutting et al., 2009; Daly et al., 2005; Hosp & MacConnell, 2008; Kim, 2015; Veenendaal, Groen, and Verhoeven, 2015). Veenendaal et al. (2015) examined the connection between ORF and reading comprehension. Results showed that text reading prosody explained additional variance in reading comprehension performance when decoding efficiency and language comprehension were controlled for, and natural intonation was associated with better comprehension of what was read. Rate and prosody are differently associated with reading comprehension scores. Text reading prosody explained additional variance in reading comprehension scores, after decoding efficiency and language comprehension were accounted for, but text reading rate did not (Veenendaal et al., 2015).

The addition of text reading prosody to the construct of ORF results in text reading prosody being the key factor, even after controlling for decoding efficiency and language comprehension, and text reading prosody contributed to reading comprehension scores (Veenendaal et al., 2015). Text reading prosody made a contribution to reading comprehension after an independent measure of syntactic ability was controlled for, which suggests that the extent to which children applied the correct prosody while reading contributed to comprehension above and beyond the influence of syntactic ability and size of vocabulary. A correct use of text reading prosody can either be a facilitator of reading comprehension or be a reflection of the level of reading comprehension (Veenendaal et al., 2015).

Kim (2015) also investigated the relationship between ORF and reading comprehension. The study examined children in prekindergarten at age 5 and again in kindergarten at age 6. The relationships between word-reading fluency, listening comprehension, and text-reading fluency and reading comprehension and between reading comprehension and text-reading fluency were examined, as were predictors of text-reading fluency, word-reading fluency, and reading comprehension (Kim, 2015).

Results indicated that listening comprehension was related to text-reading fluency, particularly after children developed a certain level of reading proficiency (Kim, 2015). Reading comprehension was related to text-reading fluency over and above word-reading fluency and listening comprehension. Text-reading fluency predicts reading comprehension, and reading comprehension predicts text-reading fluency. Listening comprehension was not independently related to text-reading fluency. Reading comprehension was independently related to text-reading fluency. Reading comprehension and text-reading fluency both involve decoding processes, and listening comprehension does not, which explains the independent relationship between reading comprehension and text-reading fluency. Vocabulary and grammatical knowledge were independently related to text-reading fluency and reading comprehension (Kim, 2015).

The relationship between text-reading fluency and reading comprehension changes over time (Bashir & Hook, 2009; Berninger et al., 2006; Cutting et al., 2009; Hosp & MacConnell, 2008; Kim, 2015) Kim (2015) found that in earlier reading development, word-reading fluency and text-reading fluency were independently related to reading comprehension after accounting for listening comprehension, but in later reading development, only text-reading fluency, but not word-reading fluency, was independently related to reading comprehension (Kim, 2015).

Catts et al. (2015) examined early precursors of reading comprehension using assessments of word reading, letter knowledge, phonological awareness, rapid naming, and oral language for 366 participants in the beginning of kindergarten and reading comprehension at the end of third grade. Results indicated that word reading precursors were moderately related to each other. The strongest relationship was found between phonological awareness and letter knowledge (Catts et al., 2015).

In addition, phonological awareness and letter knowledge were at least partially independent, accounted for variance in word recognition, and were separate constructs (Catts et al., 2015). Word-reading precursors and oral language ability were associated at the beginning of kindergarten, and the strongest relationship was found between phonological awareness and oral language. Oral language and rapid naming had a direct association with reading comprehension in third grade. Letter knowledge and phonological awareness were found to be indirectly related to reading comprehension through their unique associations with second grade word reading ability. The strongest predictor was oral language, followed by phonological awareness and rapid naming, which indicates that word reading precursors are related to later reading comprehension (Catts et al., 2015).

Reading comprehension is influenced by metacognition (Tarchi, 2015; Zabrucky et al., 2015). Zabrucky et al. (2015) examined how the components of declarative knowledge of assessment and strategy were related to the comprehension of expository texts on a comprehension exam, using an expanded metacomprehension scale. The expanded scale included measures of evaluation or awareness of comprehension during reading, regulation of comprehension, and strategies used during reading. The study found that the students' self-assessments of evaluation and regulation were related to comprehension performance. Students who reported being aware of their level of understanding regulated their reading by adjusting to difficult material. These students also identified and made connections across main points using explanatory strategies and performed better on a comprehension test. Those students who reported relying more on the use of external aid strategies performed more poorly. These results indicate a need for teachers to teach students how to use comprehension strategies (Zabrucky et al., 2015).

One of the core components of reading comprehension is inferential processes (Kendeou et al., 2016; Tarchi, 2015). Reading comprehension involves the construction of a coherent mental representation of the text in the reader's memory through inference making. A student is required to make an inference when integrating information provided by the text but found in different locations or when incorporating information obtained outside of the text (Tarchi, 2015). An inference is information that is retrieved from memory or generated during reading to fill in information that is not in a text, and the ability to make inferences is one of the unique, significant predictors of reading comprehension (Kendeou et al., 2016). An inference allows a student to integrate multiple sentences of text in order to comprehend a larger meaning of the text (Tarchi, 2015).

Evidence shows that inference-making skills develop before formal reading instruction begins, and the development of successful inference skills usually prevents later comprehension difficulties (Kendeou et al., 2016; Tarchi, 2015). The ability to formulate both constructive and semantic inferences contributes to successful reading comprehension. A constructive inference refers to creating links within the text, and a semantic inference is based on understanding the meaning of a word based on context (Tarchi, 2015). Inference-making to construct a mental representation of the content of the text is the process of reading comprehension, and the mental representation is the product of reading comprehension (Kendeou et al., 2016).

The reader's knowledge is another factor that is important for reading comprehension (Kendeou et al., 2016; Tarchi, 2015). At different levels of the reading comprehension process, the reader draws on different sources of knowledge. Prior knowledge encompasses an individual's multidimensional and hierarchical knowledge base and may include declarative knowledge, which refers to knowledge of facts and concepts, and procedural knowledge, which is knowledge related to how to carry out tasks (Tarchi, 2015). Sources of knowledge include linguistic knowledge, orthographic knowledge, and general knowledge. Accurate knowledge can facilitate reading comprehension, whereas inaccurate knowledge can disrupt reading comprehension. Readers with inaccurate knowledge generate incorrect inferences during reading, which results in impoverished mental representations of the text content (Kendeou et al., 2016). Topic knowledge of facts and meanings and understanding the meaning of specific concepts influence reading comprehension (Tarchi, 2015).

Successful reading comprehension also depends on the construction of a coherent representation of text in memory (McMaster et al., 2015). A coherent representation includes important information in the text that is integrated with the reader's background knowledge and can be easily accessed and applied. While reading, a student makes connections among important parts of text, which builds structure and coherence

(McMaster et al, 2015). Reading comprehension requires the reader to have and access relevant background knowledge, make connections among information from the text, and integrate background knowledge with text-based information to make inferences (Kendeou et al., 2016; McMaster et al., 2015; Tarchi, 2015). Inference-making is needed to construct a coherent representation.

A multitude of factors influence the complex higher order domain of reading comprehension in intricate ways (Kendeou et al., 2016; Russo et al., 2009; Tarchi, 2015; Zabrucky et al., 2015). It is important to consider the various skills required for successful reading comprehension as well as the interaction of these areas. Efforts to prevent reading comprehension difficulties are necessary and should consider frameworks that specify components, such as inferential processes and prior knowledge, that provide a basis for developing assessments and instructional approaches focused on improving reading skills.

Measures. Reading comprehension assessments typically require a student to read sentences and short passages and respond orally to questions concerning the meaning of the text (McCloskey & Perkins, 2013). Reading comprehension may be assessed with a variety of methods, including retelling, maze or cloze measures, or answering comprehension questions. When using retelling as a measure of assessment, the student states what is remembered about the content orally, and retelling responses are scored based on the number of words recalled or on a coding system (Desoff, 2007; University of Oregon Center on Teaching and Learning, 2009). Student memory may influence performance on this assessment measure. When using maze measures, the student reads sentences with blanks and fills in the missing words (Marcotte & Hintze, 2009). Standardized assessments that include a maze measure are the Passage Comprehension section of the Woodcock-Johnson-III Tests of Achievement (Woodcock, Mather, & McGrew, 2001) and the Woodcock Diagnostic Reading Battery (Woodcock, Mather, & Schrank, 2005).

When using comprehension questions as an assessment measure, the student answers open- or closed- ended questions concerning previously read material (Reutzel & Hollingsworth, 1993). The questions may be in multiple choice, fill-in-the-blanks, oral, or written format (Wise et al., 2010). Using this format on standardized reading assessments administered individually may be time consuming. Individually administered comprehension assessments include the Wechsler Individual Achievement Test – third edition (Wechsler, 2009), the Kaufman Test of Educational Achievement – third edition (Kaufman & Kaufman, 2014), the Gray Oral Reading Test – fifth edition (Widerholt & Bryant, 2012), Group Reading Assessment and Diagnostic Evaluation (Williams, 2001), Test of Reading Comprehension – fourth edition (Brown, Hammill, & Wiederholt, 2009), Developmental Reading Assessment – second edition PLUS (Beaver & Carter, 2011), and running records (Herbert, 2004).

The Daze component of the DIBELS is a measure of reading comprehension (University of Oregon Center on Teaching and Learning, 2018). Daze is a group-administered maze measure where students are instructed to read a passage silently. In the passage, every seventh word is blank, and students are given three possible word choices for the blank. Students are required to choose the correct word as they read the passage. Students are given 3 minutes to complete the Daze task (University of Oregon Center on Teaching and Learning, 2018).

Interventions. Reading comprehension is a necessary component for student achievement in reading, and implementing appropriate interventions is important to student success. Students who successfully learn to decode text in early elementary grades may struggle in later grades, when requirements become increasingly complex (McMaster et al., 2015). Problems with comprehension skills, which usually increase with age as reading material becomes more challenging, should be addressed with additional instruction and interventions (Fletcher et al., 2011). Reading comprehension difficulties that were previously unnoticed and untreated can emerge in later elementary school years. For college and career readiness, students are expected to be able to read and comprehend a range of texts across content areas independently and proficiently (Ritchey, Palombo, Silverman, & Speece, 2017).

Combining multiple strategies and interventions may increase student success in reading comprehension (Daly et al., 2005; NICHHD, 2000; University of Oregon Center on Teaching and Learning, 2009). Strategies used before, during, and after reading include prereading, rereading, previewing, analyzing purpose, slowing reading speed, revisiting text, reading aloud, outlining, taking notes, analyzing story structure, chunking, activating background knowledge, connecting to other text, modeling, questioning, content discussion, and summarizing. Other strategies include the use of graphic and semantic organizers, flow charts, pictorial aids, and scaffolding. Students should be exposed to various text genres through innovative and effective use of technology and different media (Kendeou et al., 2016). Key factors of effective interventions include explicit instruction in reading comprehension strategies, such as main idea identification, summarization, question answering, attention to self-regulation, and peer collaboration (Ritchey et al., 2017).

Comprehension Plus and The Comprehension Toolkit are interventions that provide instruction in reading comprehension strategies. Comprehension Plus is a program that teaches strategies through lessons delivered twice a week to small or large groups of students (Pearson Education, Inc., 2018). The Comprehension Toolkit includes lessons that help students learn how to understand, respond to, and learn from text (Harvey & Goudvis, 2018). Strategies are taught through instruction, modeling, and guided practice. Strategies include connecting, asking questions, inferring meaning, determining importance, summarizing, and synthesizing.

Interventions that address reading comprehension as well as other components of reading include Project Read, Alphabetic Phonics, Explode The Code, and Reading Milestones. Project Read is designed to systematically teach skills through multisensory activities and practice exercises (Language Circle Enterprises, 2018). Students learn how to integrate decoding, vocabulary development, narrative and expository reading, and questioning strategies to foster reading comprehension. Alphabetic Phonics teaches reading comprehension in addition to phonics, language structure, handwriting, and spelling through a multisensory curriculum (EPS School Specialty, 2018). Explode The Code is designed to provide a systematic approach in online or workbook formats to teach comprehension skills in addition to other reading skills (EPS School Specialty, 2018). Reading Milestones addresses vocabulary and comprehension skills, including literal, inferential, evaluative, and critical thinking skills (PRO-ED, 2018).

Research shows that it is helpful for interventions to address the higher order skills required for reading comprehension (Kendeou et al., 2016; Tarchi, 2015; Zabrucky et al., 2015). Tarchi (2015) examined the efficacy of a reading comprehension program that was based on the activation of relevant dimensions of prior knowledge, as well as student metacognition. Students learned how to control their reading comprehension process. The intervention had a positive effect on student reading comprehension, metacognitive approach to reading, and inference-making processes. Discussing how prior knowledge relates to the text was helpful. The focus on multiple dimensions of prior knowledge helped students to improve their semantic inference-making skills (Tarchi, 2015).

McMaster et al. (2015) also explored an intervention related to higher level components of reading comprehension. This study examined the effect of questioning to facilitate coherent representations as an intervention for reading comprehension. Questioning helps to support connections between parts of text and constructing coherent representations. General and causal questions and question timing were explored. Causal questions refer to questions that direct the reader's attention to causal relations and help the reader to identify logical relations among events in the text. Questions were asked during reading and after reading. Questions asked during reading allow for emphasis on cognitive processes that operate during reading, but may interrupt the reading process and constrain attention and working memory (McMaster et al., 2015).

Results indicated that causal questions asked during or after reading were more helpful in forming coherent representations than asking general questions after reading (McMaster et al., 2015). The intervention was implemented with high fidelity and is feasible in educational settings. Prompting is important to supporting student identification of text-based connections. Prompting and asking causal questions is a promising intervention to address the forming of coherent representations in the process of reading comprehension (McMaster et al., 2015).

Ritchey et al. (2017) explored the efficacy of a short-term informational text reading comprehension intervention that focused on reading comprehension strategy instruction. The intervention included explicit instruction, practice in authentic texts such as trade books, and peer interaction. Results indicated that the intervention was successful in teaching students the targeted strategies. In addition, the intervention is feasible in an RTI framework because the recommended timeframe of the intervention is 10 to 12 weeks. This intervention holds promise for use as a Tier 2 or Tier 3 intervention (Ritchey et al., 2017).

It is important for interventions to take into account the various skills required for reading comprehension and the interaction of these skills (Kendeou et al., 2016; Tarchi, 2015; Zabrucky et al., 2015). Intervention implementation should be flexible and consider basic reading skills, such as phonemic awareness, decoding, sight word recognition, fluency, accuracy, and vocabulary, as well as higher level components, such as inference-making, access to background knowledge, understanding explicit information, inferential comprehension, evaluative and critical reading, and integration of information. Similarly to ORF, reading comprehension interventions should be implemented in consideration of executive functions, strategy use, and in the context of the other components of the Balanced Literacy model (Lennon, 2017; McCloskey, 2016). It is helpful for students to be aware of their reading comprehension abilities and how these abilities fit into the larger framework of the reading process and learn how to use strategies related to interventions before using these strategies.

Curriculum-based Measurement and Statewide Testing

Research has explored the connection between reading CBM and statewide testing of reading performance. The amount of research linking CBM and performance on statewide assessments has continued to increase, and CBM can be considered a useful tool in predicting outcomes on statewide achievement tests in reading (Shapiro et al., 2006; Wood, 2006). Performance on CBM may also be useful in identifying students who are at risk for failing state tests and informing educators of the need for intervention (Barger, 2003; Shaw & Shaw, 2002; Stage & Jacobsen, 2001). Students who are at risk of failing state testing may receive tailored instruction early in the school year and be referred for further problem analysis of the skill deficits in order to increase the chance of passing (Stage & Jacobsen, 2001).

Research shows that there is a significant relationship between high reading ability and successive school performance (Woodward & Talbert-Johnson, 2009). Early literacy performance continues to be related to later reading performance, and the DIBELS measures seem to be reliable indicators of reading achievement in subsequent years (Utchell, Schmitt, McCallum, McGoey, & Piselli, 2016). Research suggests that the role and strength of specific early literacy measures can change over time. As reading competency develops, alphabetic principle measures such as Letter Naming Fluency and Nonsense Word Fluency tend to retain predictive value, whereas phonological awareness measures such as Phoneme Segmentation Fluency lose predictive strength. Research has yet to firmly establish the ability of ORF to predict state assessment performance beyond

2 years and the ability of kindergarten early literacy measures to add unique information in predicting future statewide test performance (Utchell et al., 2016).

Research has examined the connection between ORF CBM and statewide tests. In a multiyear study using a large sample size with a large percentage of low socioeconomic status and non-Caucasian students, McGlinchey and Hixson (2004) found that ORF is a good predictor of performance on state reading assessments. Materials included a CBM probe consisting of reading passages to measure ORF and the Michigan Educational Assessment Program, a fourth grade statewide reading assessment. This study found that CBM can be used to identify students at risk for failing state-mandated reading assessments. Reading CBM may also help educators to make instructional decisions and adjustments and provide for prereferral and referral processes and IEP progress monitoring (McGlinchey & Hixson, 2004).

The increased accountability of schools to produce competent readers has led to increased interest in establishing the relationship between DORF and statewide test scores. There is a consensus in the literature that within a school year, ORF and state reading assessment performances are highly correlated, and there is evidence of long-term predictive validity of ORF (Shapiro et al., 2006, 2008; Utchell et al., 2016). Many researchers have demonstrated the crucial importance of specific emergent literacy skills for later successful reading development. In order to help students with emergent literacy deficits catch up to typically developing peers, empirically validated early instruction and intervention approaches are necessary (Utchell et al., 2016). ORF measures have been identified as accurate predictors of performance on statewide end-of-year reading assessments (Goffreda et al., 2009).

Longitudinal studies of kindergarten students have shown that early literacy skills are predictive of performance on individually administered standardized reading measures in first and second grade and that literacy measures such as the DORF have predictive validity. Research has found that DORF is predictive of student performance on the PSSA (Shapiro et al., 2006, 2008; Utchell et al., 2016).

The PSSA is a criterion-referenced measure based on a standards-aligned system matched with state determined grade-level standards to assess student proficiency in reading, math, science, and writing content areas (Pennsylvania Department of Education [PDE], 2016, 2018). It is administered in all public schools in Pennsylvania to students in Grades 3 through 8. This assessment provides educational accountability information and determines if schools make adequate yearly progress by providing data on student achievement of specified standards. The PSSA offers an alternate assessment for students with severe cognitive disabilities, the Pennsylvania Alternate System of Assessment. This alternate form is provided for students with Individualized Education Plans and Section 504 Service Agreement plans (PDE, 2018).

The PSSA yields numeric descriptions that are performance level scores. For each grade and subject, there are three cut scores that distinguish between performance levels. The four performance levels are: Level 1 – Below Basic, Level 2 – Basic, Level 3 – Proficient, and Level 4 – Advanced (Northwest Evaluation Association [NWEA], 2016; PDE, 2018). The Level 3 cut score defines the minimum level of performance considered to be Proficient. A score of Advanced indicates superior skills, a score of Proficient indicates satisfactory skills, a score of Basic indicates limited skills, and a score of Below Basic indicates inadequate skills (PDE, 2016). The goal of NCLB was for all students to demonstrate Proficient or Advanced performance compared to grade-level standards.

Utchell et al. (2016) explored the predictive efficiency of ORF CBM for performance on the PSSA. This study examined the extent to which early literacy measures administered in kindergarten and ORF measures administered in Grade 1 are related to and predict future state reading assessment performances up to 7 years later. This study explored the relationships among kindergarten DIBELS Initial Sound Fluency, Phoneme Segmentation Fluency, Letter Naming Fluency, and Nonsense Word Fluency, Grade 1 ORF, and PSSA performance in Grades 3, 5, and 7. It also examined the extent to which the early literacy measures administered at these points predict future state assessment scores up to 7 years later in Grades 3, 5, and 7 and whether the measures predict future achievement, while controlling for previous state performance (Utchell et al., 2016).

The participants in this study were a district-wide group of kindergarten students in a suburban school district in southwestern Pennsylvania (Utchell et al., 2016). The study utilized existing data available as a result of the universal screening efforts of the participating school district and included the data points of kindergarten DIBELS, Grade 1 ORF, and Grades 3, 5, and 7 state assessment scores. ORF was a significant predictor of state reading performance in Grade 3. DIBELS measures accounted for additional significant variance in state assessment performance. Early literacy measures could be used to predict performance on a standardized reading comprehension test 2 years later, and the results add further evidence of the predictive validity of DIBELS measures well beyond 2 years in the future. ORF significantly predicted performance above and beyond previous state assessments on state standardized assessments up to 5 years later (Utchell et al., 2016).

Shapiro et al. (2006) examined two school districts in Pennsylvania and found that CBM reading measures had moderate to strong relationships with PSSA performance. CBM measures obtained during the winter or spring assessment periods were strong predictors of subsequent high-stakes achievement assessment, whether the student was below or above the criterion on that measure (Shapiro et al., 2006). Schools have used DORF to predict the possibility of students passing statewide assessment, and research shows a positive correlation between DORF scores and statewide assessment scores (Wanzek & Vaughn, 2011). DORF, as compared to other DIBELS measures, is a significant predictor of PSSA reading proficiency levels (Goffreda et al., 2009).

More recently, Shapiro et al. (2008) similarly found that DORF predicted PSSA scores. The combination of DORF with an additional reading benchmark assessment provided the best predictive outcomes. This study explored Grades 3 through 5 using data from fall and winter assessments of DORF and the 4Sight Benchmark Assessment, along with PSSA results (Shapiro et al., 2008).

The association between DIBELS ORF and Daze scores with statewide testing results has also been explored. Kim, Vanderwood, and Lee (2016) examined the predictive validity and accuracy of CBM in reading for third grade Spanish-speaking English learners at various levels of English proficiency, using DIBELS ORF and Daze scores. Both ORF and Daze are significant predictors of reading outcomes for English Learners. Results revealed that ORF accounted for more variance and was a stronger predictor of reading outcomes than Daze. Daze was a significant predictor when examined individually, but did not explain significant additional variance beyond ORF. There was not a significant difference in the predictive validity of ORF or Daze for students of varying English proficiency levels, and the predictive accuracy of ORF and Daze cut-scores varied by English proficiency levels (Kim et al., 2016).

Munger, LoFaro, Kawryga, Sovocool, and Medina (2014) examined the validity of the DIBELS ORF and Daze in a sample of 85 third and fifth grade students. Tests administered included DIBELS, Peabody Picture Vocabulary Test – IV, Group Reading Assessment and Diagnostic Evaluation (GRADE), and the New York State English Language Arts (NYSELA) test. Results revealed that DIBELS ORF and Daze subtest scores were significantly correlated with GRADE and NYSELA scores. Daze scores explained significant variance in GRADE scores beyond ORF at the third grade level, but did not explain variance in NYSELA scores in either grade level (Munger et al., 2014).

DIBELS ORF scores were found to be strongly and significantly correlated with reading comprehension scores from both the GRADE and the NYSELA (Munger et al., 2014). The ORF scores were a powerful predictor of New York State test scores, even when compared to a measure of considerable length and psychometric stability, such as GRADE. ORF scores were found to be strongly and consistently related to Daze scores across grade levels, and ORF was strongly related to reading comprehension scores (Munger et al., 2014).

DIBELS Daze was also found to be strongly and significantly correlated with multiple measures of comprehension, including NYSELA scores (Munger et al., 2014). Daze scores did not reliably explain variance in reading comprehension across different instruments and grade levels. Variance in reading comprehension unaccounted for by

ORF was accounted for by Daze at certain grade levels. DIBELS ORF scores do share notable variance with reading comprehension, but a measure of language comprehension could account for additional variance. DIBELS ORF and Daze may be measuring similar underlying constructs of reading such as word recognition (Munger et al., 2014).

Research has also examined the predictive validity of DIBELS Total score (Ferchalk, 2013; Ferchalk, Cogan-Ferchalk, & Richardson, 2012; Good et al., 2011; Prosser, 2015). DIBELS Total score is able to accurately classify students as having or not having a potential reading problem (Prosser, 2015). Ferchalk et al. (2012) assessed the correlations between DIBELS ORF, Daze, Total score, Retell Fluency, and Reading Accuracy with the PSSA. Scores for 184 third grade students on the DIBELS in the fall, winter, and spring and the PSSA in the spring were assessed. The strongest correlation was found between the ORF and the PSSA, followed by the DAZE. The DIBELS Total score outperformed ORF and the individual DIBELS indicators. None of the researcher-generated composite scores demonstrated stronger correlations with the PSSA than the DIBELS Total score (Ferchalk et al., 2012).

Although DIBELS Total score may provide a more reliable overall measure of general reading outcomes than DIBELS individual measures, research does not support the necessity to administer all DIBELS measures to obtain a Total score (Dynamic Measurement Group, 2010; Ferchalk et al., 2012; Good et al., 2011; University of Oregon Center on Teaching and Learning, 2012, 2013). The Total score is not a consistently better predictor of student performance than DIBELS individual measures. The Total score does not allow for consideration of each foundational skill in isolation (University of Oregon Center on Teaching and Learning, 2013).

Research supports the use of CBM in predicting statewide test performance (NWEA, 2016; Shapiro et al., 2006, 2008; Utchell et al., 2016). This allows for identification of students who are at risk for failing assessments and providing appropriate intervention. As a result of high-stakes assessments, there is a necessity to continue to determine effective means of progress monitoring and the predictive efficiency of CBM for statewide test performance. There is increased pressure on schools to produce higher scores on statewide tests due to national legislation (Bursuck & Blanks, 2010). It is necessary to explore the use of appropriate CBM methods of monitoring student progress, identifying at-risk students, and providing effective interventions to increase student performance (McGlinchey & Hixson, 2004; NWEA, 2016).

Research has yet to fully explore how reading CBM assessments can best be utilized within the RTI model (NWEA, 2016; Utchell et al., 2016). Exploring the accuracy of CBM in assessing student reading performance and the predictive efficiency of CBM for PSSA performance will add to knowledge concerning early identification of at-risk students. Research should continue to aim to help educators efficiently predict student performance on statewide tests as early as possible. Accurate use of CBM tools will help to improve efforts to increase student reading performance.

Current Study

The current study aimed to explore the relationship between DIBELS ORF, Daze, and Total score and PSSA performance. This study examined multiple components of the same assessment, the DIBELS, and how these components related to PSSA performance. The study added to previous research concerning CBM and statewide testing performance by providing further knowledge on how CBMs can be used most effectively. This study also focused on student performance over time to expand on prior knowledge concerning prediction of student performance on statewide tests as early as possible and the adjustment of intervention intensity over time.

Research Questions

The following research questions were addressed in this study:

Is there a significant difference between the proportion of students in the sample identified as Not Proficient on the PSSA reading assessment and the proportion of the population of students in the state of Pennsylvania identified as Not Proficient on the PSSA reading assessment, based on (a) fourth grade PSSA results and (b) fifth grade PSSA results?

What proportion of students were identified as at risk of earning a Not Proficient category rating on the PSSA reading assessment, based on (a) fourth grade fall and spring DIBELS ORF, Daze, and Total category scores and (b) fifth grade fall and spring DIBELS ORF, Daze, and Total category scores?

Are there statistically significant differences between the number of students identified as at risk on the DIBELS ORF, Daze, and Total category scores in the fall and spring, based on (a) fourth grade scores and (b) fifth grade scores?

What is the relationship between DIBELS ORF, Daze, and Total category scores (At Risk/Not At Risk) and PSSA reading assessment category scores (Proficient/Not Proficient), based on (a) fourth grade DIBELS and fourth grade PSSA scores, (b) fifth grade DIBELS and fifth grade PSSA scores, and (c) fourth grade DIBELS and fifth grade PSSA scores?

What types of DIBELS ORF, Daze, and Total category score change patterns were exhibited?

Summary of Research Questions

These research questions were intended to assess the relationship between DIBELS ORF, Daze, and Total scores in a number of ways. They examined the extent to which each DIBELS score accurately identifies students at risk for failure on the PSSA and the relative level of overestimation or underestimation of those students using the DIBELS ORF, Daze, and Total scores. This study also examined DIBELS fall and spring scores and the prediction of PSSA category scores within the fourth and fifth grades and across the fourth and fifth grades.

CHAPTER 3

METHODOLOGY

Overview

The current study analyzed archival data from a suburban-rural elementary school located in eastern Pennsylvania. Data included DIBELS ORF, Daze, and Total scores and PSSA reading scores.

Source of Data

Archived test scores of male and female students who attended a suburban-rural elementary school in eastern Pennsylvania were used in this study. The district consists of a population that is 44.7% White, 22.5% Black, and 27.1% Hispanic. In the district, 19% of the students receive special education services (Pennsylvania State Data Center, 2018). Data was collected for those students with available DIBELS and PSSA reading scores enrolled in the fourth grade during the 2015-2016 school year and subsequently enrolled in the fifth grade during the 2016-2017 school year. Data for students who did not have DIBELS ORF, Daze, Total scores, and/or PSSA scores on file were excluded. Participants included 75 students.

Materials

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The DIBELS is a CBM reading tool, consisting of different measures of early literacy skills, including First Sound Fluency, Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Retell Fluency, Oral Reading Fluency, and Daze (University of Oregon Center on Teaching and Learning, 2018). The current study focuses on ORF, Daze, and Total score of the DIBELS Next. Scores on the DIBELS measures are compared to benchmark goals and cut points for risk to yield risk status classifications. DIBELS scoring uses three cut points for risk that correspond to student chance of achieving literacy goals: At or Above Benchmark – 80-90%, Below Benchmark – 40-60%, and Well Below Benchmark – 10-20% (Dynamic Measurement Group, 2010).

DIBELS Oral Reading Fluency (ORF). DIBELS ORF is a standardized, individually administered test of accuracy and fluency with connected text (Good & Kaminski, 2002.) It is a standardized set of passages and administration procedures that is intended to assist with identifying students who may need additional reading instructional support. The ORF passages are designed to monitor progress toward instructional goals.

DIBELS ORF administration requires the student to read three grade level calibrated passages aloud for 1 minute each. Errors are coded and include omitted words, substituted words, and hesitations of more than 3 seconds. If a word is self-corrected within 3 seconds, it is considered correct. The student’s score, or ORF rate, is the median number of correct words per minute read aloud (Good & Kaminski, 2002).

Research has found that the DIBELS is a reliable and valid tool for reading CBM (Elliot, Lee, & Tollefson, 2001; Roberts et al., 2005). Research indicates that the DIBELS measures provide a reliable and valid indicator of children’s progress toward the acquisition of early literacy skills (Good & Kaminski, 2002). Various studies have indicated well-established reliability and validity of the DIBELS as a measure of ORF (Good, Gruba, & Kaminski, 2001; Good, Simmons, & Kame’enui, 2001; University of Oregon Center on Teaching and Learning, 2009). The ORF test-retest reliability ranges from .92 to .97, alternate form reliability ranges from .89 to .94, and criterion-related validity between the ORF and reading comprehension tests ranges from .60 to .90 (Shaw & Shaw, 2002).

DIBELS Daze. DIBELS Daze is a standardized measure of reading comprehension that requires students to read a passage silently and select the correct word from a list of three choices to fill in blanks occurring approximately every seventh word. Alternate forms, test-retest, and interrater coefficients are above .90, except for an alternate forms reliability coefficient for fifth grade at .74 (Good et al., 2011).

DIBELS Total Score. DIBELS Total score, also called composite score, is a combination of the DIBELS measures and provides an overall estimate of literacy skills and/or reading proficiency. All of the DIBELS measures are aggregated to form the total score. Alternate forms reliability is .66 and interrater reliability is .97 (Dynamic Measurement Group, 2010; Good et al., 2011; University of Oregon Center on Teaching and Learning, 2012).

Pennsylvania System of School Assessment (PSSA). The PSSA is a standards-based assessment that contains content-specific assessments in English Language Arts, Mathematics, and Science. The current study focuses on the English Language Arts portion of the PSSA. The PSSA is administered in the spring in all public schools within the state of Pennsylvania for students in Grades 3 through 8 (PDE, 2018).

For each grade and subject, there are three cut scores that distinguish between performance levels. The four performance levels are: Level 1 – Below Basic, Level 2 – Basic, Level 3 – Proficient, and Level 4 – Advanced (NWEA, 2016; PDE, 2018). The Level 3 cut score defines the minimum level of performance considered to be Proficient with regard to accountability.

The Below Basic level reflects inadequate academic performance. Below Basic level work indicates little understanding and minimal display of the skills included in the Pennsylvania Academic Content Standards. The Basic level reflects marginal academic performance. Basic level work indicates a partial understanding and limited display of the skills included in the Pennsylvania Academic Content Standards. This work is approaching satisfactory performance. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level. Proficient work indicates a solid understanding and adequate display of the skills included in the Pennsylvania Academic Content Standards. The Proficient level reflects satisfactory academic performance. The Advanced level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Pennsylvania Academic Content Standards (PDE, 2018).

Procedure

The current study's quantitative research design is modeled after previous studies examining the relationship between CBM and statewide testing scores (Barger, 2003; Buck & Torgesen, 2003; Crawford et al., 2001; Fuchs et al., 2001; Good et al., 2001, 2002; McGlinchey & Hixson, 2004; Roehrig et al., 2008; Schilling et al., 2007; Shapiro et al., 2006, 2008; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Wiley & Deno, 2005; Wilson, 2005; Wood, 2006). In the current study, student names were removed from the data file and replaced with identification numbers to ensure confidentiality. Demographic data in the student files included age and gender.

DIBELS scores for fall and spring of Grade 4 and Grade 5 were obtained and collapsed into dichotomous categorical scores. DIBELS scores were converted into the categories of Not Proficient (a combination of Below Benchmark and Well Below Benchmark) and Proficient (At or Above Benchmark). PSSA English Language Arts scores for Grade 4 and Grade 5 were obtained and collapsed into the categories of Not Proficient (a combination of Below Basic and Basic) and Proficient (a combination of Proficient and Advanced).

Statistical Analyses

Nonparametric descriptive statistics tests were used. Descriptive analyses were completed to examine the relationships between type of CBM and timing of CBM. Fisher's exact tests were used to examine the differences in proportions when comparing DIBELS ORF, Daze, and Total scores and when comparing fall and spring administrations. Cross-tabulation table values were used to calculate values for Improvement, Instability, Sensitivity, Specificity, and Kappa indices and percentage of students determined to be at risk. Values were derived from the generation of 2 x 2 cross-tabulation tables using the formulas shown in Table 1.

Table 1

Construction of Cross-tabulation Tables and Calculation Formulas Used to Derive the Index Values Used in Statistical Analyses of Data

DIBELS Fall or Spring Score	PSSA Score Category	
Category	Not Proficient	Proficient
Not Proficient	A	B
Proficient	C	D

Improvement Index = (B/(A+B)) x 100

Instability Index = (C/(C+D) x 100

Sensitivity Index = (A/(A+C)) x 100

Specificity Index = (D/(B+D)) x 100

Kappa Index = ((po-pe)/(1-e)) x 100 where:

Po = pA +pD

Pe = ((pA +pC)(pA+pB)) + ((pB +pD)(pC+pD))

pA=A/Total N pB=B/Total N pC=C/Total N pD=D/Total

Operational definitions for the indices and patterns used to analyze the data and interpret findings were as follows:

At Risk: A student was deemed to be at risk of failing to obtain a Proficient score on the PSSA if the student obtained a Not Proficient score on either the fall or spring administration of DIBELS.

Percentage of Students At Risk: The percent of students at risk was operationally defined as the percentage of students at risk of not being Proficient on the PSSA, based on the results of a DIBELS administration.

Improvement Index: The Improvement Index was operationally defined as the percentage of students categorized as Not Proficient on a DIBELS administration, but identified as Proficient on the PSSA. The Improvement Index represents the success rate of students identified as at risk of being Not Proficient on the PSSA.

Instability Index: The Instability Index was operationally defined as students who were identified as Proficient on a DIBELS administration, but who conversely earned scores in the Not Proficient range on the PSSA.

Sensitivity Index: Sensitivity was operationally defined as the proportion of students who were identified as Not Proficient on the PSSA and were also identified as Not Proficient on a DIBELS administration.

Specificity Index: Specificity was operationally defined as the proportion of students who were identified as Proficient on the PSSA and were also identified as Proficient on a DIBELS administration.

Kappa Index: The Kappa statistic indicates the percentage of increase over chance represented by the overall percentage of agreement of DIBELS categories with PSSA results.

Fisher's exact tests were used to examine the differences in proportions obtained for the various indexes when comparing DIBELS ORF, Daze, and Total scores, when comparing scores within and across fall and spring, and when comparing scores across Grades 4 and 5.

Analysis of Performance Patterns Using DIBELS and PSSA Score Categories: In addition to the calculation and statistical analysis of the index scores, a descriptive analysis was completed to examine the relationship between DIBELS fall and spring ORF, Daze, and Total scores and PSSA results. To accomplish this descriptive analysis, performance patterns involving the fall and spring administrations of the DIBELS ORF, Daze, and Total scores and PSSA results were identified.

A performance pattern was determined for each student by examining the category scores obtained on the fall and spring administrations of the DIBELS ORF, Daze, and Total scores and the spring administration of the PSSA and categorizing patterns of changes in status from fall DIBELS scores to spring DIBELS scores to spring

PSSA scores. Students were assigned to categories based on the pattern of relationship among these three scores. Percentages of students exhibiting each performance pattern were calculated and placed in a table for descriptive analysis. This procedure was used to compare fourth grade DIBELS scores with fourth grade PSSA results, fourth grade DIBELS scores with fifth grade PSSA results, and fourth grade DIBELS scores with fifth grade PSSA results.

In the table, patterns were designated with the letters N for Not Proficient or At Risk and P for Proficient or Not At Risk. Using this coding, the following patterns were identified:

Consistently Negative Pattern:

N-N-N = DIBELS fall score at risk, DIBELS spring score at risk, PSSA not proficient

Negative Change Patterns:

P-N-N = DIBELS fall score not at risk, DIBELS spring score at risk, PSSA not proficient

P-P-N = DIBELS fall score not at risk, DIBELS spring score not at risk, PSSA not proficient

N-P-N = DIBELS fall score at risk, DIBELS spring score not at risk, PSSA not proficient

Positive Change Patterns:

P-N-P = DIBELS fall score not at risk, DIBELS spring score at risk, PSSA proficient

N-N-P = DIBELS fall score at risk, DIBELS spring score at risk, PSSA proficient

N-P-P = DIBELS fall score at risk, DIBELS spring not at risk, PSSA proficient

Consistently Positive Pattern:

P-P-P = DIBELS fall not at risk, DIBELS spring not at risk, PSSA proficient

CHAPTER 4

RESULTS

It was predicted that the percentage of students who failed the PSSA would not significantly differ from statewide results. Fisher's exact tests were conducted to compare (a) the sample's fourth grade PSSA reading assessment scores to the population of students in the state of Pennsylvania identified as Not Proficient on the fourth grade PSSA administered in the spring of the 2015-16 school year and (b) the sample's fifth grade PSSA reading assessment scores to the population of students in the state of Pennsylvania identified as Not Proficient on the fifth grade PSSA administered in the spring of the 2016-17 school year.

Results indicated that the difference between groups for the fourth grade PSSA of the 2015-16 school year was not significant ($p = 0.244$). Statewide, 41% of students earned scores in the Not Proficient range. In this study, 34.7 % earned scores in the Not Proficient range. The difference between groups for the fifth grade PSSA of the 2016-2017 school year was also not significant ($p = 0.212$). For the fifth grade PSSA of the 2016-17 school year, 40% of students statewide earned scores in the Not Proficient range, similar to this study's 33.3% of students earning scores in the Not Proficient range (Pennsylvania Department of Education, 2017).

As shown in Table 2, of the 75 fourth grade students who took the ORF in the fall, 36 (48%) were identified as at risk of earning a Not Proficient category rating on the PSSA reading assessment. In the spring, the number of students decreased slightly to 35 (47%). On the Daze, in the fall, 29 (39%) were identified as at risk and 31 (41%) in the spring. For the Total category, 25 (33%) were identified as at risk in the fall, and 29 (39%) were identified as at risk in the spring.

Table 2

Percentage of Students Identified as At Risk of Not Passing PSSA Based on DIBELS Category Scores

	Percent at risk			
	Based on Grade 4 DIBELS category score		Based on Grade 5 DIBELS category score	
DIBELS	Fall	Spring	Fall	Spring
ORF	48	47	49	51
Daze	39	41	47	43
Total	33	39	45	45

Fisher's exact tests were conducted to compare the fall to the spring administrations regarding number of students identified as at risk on the DIBELS ORF, Daze, and Total category scores. These values are provided in Table 3. There were no statistically significant differences between fall and spring administrations. Fisher's exact tests also showed that there were no statistically significant differences among the numbers of students identified as at risk between ORF, Daze, and Total category scores within fall and spring.

As shown in Table 2, of the 75 fifth grade students who took the ORF in the fall, 37 (49%) were identified as at risk of earning a Not Proficient rating on the PSSA reading assessment. In the spring, 38 (51%) were identified as at risk. On the Daze, in the fall, 35 (47%) were identified as at risk, and in the spring, 32 (43%) were identified as at risk. For the Total category, 34 (45%) were identified as at risk in the fall as well as in the spring.

Table 3

Statistical Tests of Significance Comparing the Number of Fourth Grade Students Identified as At Risk Based on DIBELS Scores

	At risk based on Grade 4 DIBELS category score (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	36	35	0.164	0.869
Daze	29	31	-0.333	0.739
Total	25	29	-0.680	0.497
Fall	D1	D2		
ORF (D1) vs Daze (D2)	36	29	1.153	0.249
ORF (D1) vs Total (D2)	36	25	1.828	0.068
Daze (D1) vs Total (D2)	29	25	0.680	0.497
Spring	D1	D2		
ORF (D1) vs Daze (D2)	35	31	0.658	0.511
ORF (D1) vs Total (D2)	35	29	0.991	0.322
Daze (D1) vs Total (D2)	31	29	0.333	0.739

As shown in Table 4, Fisher's exact tests found no statistically significant differences between fall and spring administrations regarding number of students identified as at risk on the DIBELS ORF, Daze, and Total category scores. In addition, Fisher's exact tests showed that there were no statistically significant differences among the numbers of students identified as at risk on the ORF, Daze, and Total category scores within fall and spring.

Based on previous research examining the predictive efficiency of the DIBELS for statewide testing (Dynamic Measurement Group, 2010; Ferchalk et al., 2012; Good et al., 2011; Shapiro et al., 2006, 2008; University of Oregon Center on Teaching and Learning, 2012, 2013; Utchell et al., 2016), it was hypothesized that ORF, Daze, and Total categories would be predictive of PSSA results and would not significantly differ in the identification of percentage of students at risk of earning a Not Proficient category rating on the PSSA reading assessment, as well as in the values obtained for the Improvement, Instability, Sensitivity, Specificity, and Kappa indices.

The proportion of students identified as at risk with fall and spring ORF, Daze, and Total category scores who earned PSSA scores in the Proficient range is reported as the Improvement Index. These values are displayed in Table 5.

Fourth grade fall administrations of ORF, Daze, and Total categories yielded Improvement Index values of 42%, 31%, and 28%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Improvement Index values of 49%, 39%, and 34%, respectively.

Fifth grade fall administrations of ORF, Daze, and Total categories yielded Improvement Index values of 49%, 40%, and 44%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Improvement Index values of 45%, 41%, and 41%, respectively.

Table 4

Statistical Tests of Significance Comparing the Number of Fifth Grade Students Identified as At Risk Based on DIBELS Scores

	At risk based on Grade 5 DIBELS category score (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	37	38	-0.163	0.870
Daze	35	32	0.493	0.622
Total	34	34	0.000	1.000
Fall	D1	D2		
ORF (D1) vs Daze (D2)	37	35	0.327	0.744
ORF (D1) vs Total (D2)	37	34	0.491	0.623
Daze (D1) vs Total (D2)	35	34	0.164	0.870
Spring	D1	D2		
ORF (D1) vs Daze (D2)	38	32	0.982	0.326
ORF (D1) vs Total (D2)	38	34	0.654	0.513
Daze (D1) vs Total (D2)	32	34	-0.329	0.742

Table 5

Improvement Index Values Based on DIBELS and PSSA Category Scores

	Improvement Index					
	Grade 4 DIBELS and Grade 4 PSSA		Grade 5 DIBELS and Grade 5 PSSA		Grade 4 DIBELS and Grade 5 PSSA	
DIBELS	Fall	Spring	Fall	Spring	Fall	Spring
ORF	42%	49%	49%	45%	42%	43%
Daze	31%	39%	40%	41%	41%	39%
Total	28%	34%	44%	41%	32%	41%

When comparing fourth grade DIBELS category scores with fifth grade PSSA scores, fall administrations of ORF, Daze, and Total categories yielded Improvement Index values of 42%, 41%, and 32%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Improvement Index values of 43%, 39%, and 41%, respectively.

As shown in Table 6, Fisher's exact tests indicated that there were no significant differences in Improvement Index values between fourth grade fall and spring administrations of ORF, Daze, and Total score categories. In addition, all comparisons of the Improvement Index between ORF, Daze, and Total score categories within fall and spring of fourth grade were nonsignificant.

Table 6

Statistical Tests of Significance Comparing the Improvement Index Values Based on Fourth Grade DIBELS Scores and Fourth Grade PSSA Scores

	Improvement based on Grade 4 DIBELS and Grade 4 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	15	17	-0.585	0.558
Daze	9	12	-0.623	0.533
Total	7	10	-0.511	0.609
Fall	D1	D2		
ORF (D1) vs Daze (D2)	15	9	0.883	0.377
ORF (D1) vs Total (D2)	15	7	1.093	0.274
Daze (D1) vs Total (D2)	9	7	0.243	0.808
Spring	D1	D2		
ORF (D1) vs Daze (D2)	17	12	0.806	0.420
ORF (D1) vs Total (D2)	17	10	1.136	0.256
Daze (D1) vs Total (D2)	12	10	0.340	0.734

There were no significant differences in Improvement Index values between fifth grade fall and spring administrations of ORF, Daze, and Total score categories. These results are displayed in Table 7. All comparisons of the Improvement Index between

ORF, Daze, and Total score categories within fall and spring of fifth grade were nonsignificant.

Table 7

Statistical Tests of Significance Comparing the Improvement Index Values Based on Fifth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Improvement based on Grade 5 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	18	17	0.339	0.735
Daze	14	13	-0.052	0.959
Total	15	14	0.245	0.807
Fall	D1	D2		
ORF (D1) vs Daze (D2)	18	14	0.738	0.461
ORF (D1) vs Total (D2)	18	15	0.382	0.703
Daze (D1) vs Total (D2)	14	15	-0.346	0.729
Spring	D1	D2		
ORF (D1) vs Daze (D2)	17	13	0.346	0.729
ORF (D1) vs Total (D2)	17	14	0.305	0.760
Daze (D1) vs Total (D2)	13	14	-0.046	0.963

As shown in Table 8, when comparing fourth grade DIBELS category scores with fifth grade PSSA category scores, there were no significant differences between Improvement Index values for the fall and spring administrations of DIBELS category scores. In addition, there were no significant differences between Improvement Index values of DIBELS categories within fall and spring.

The proportion of students identified as Proficient with fall and spring ORF, Daze, and Total category scores and who earned PSSA scores in the Not Proficient range is reported as the Instability Index. Instability Index values are displayed in Table 9.

Fourth grade fall administrations of ORF, Daze, and Total categories yielded Instability Index values of 13%, 13%, and 16%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Instability Index values of 20%, 16%, and 15%, respectively.

Fifth grade fall administrations of ORF, Daze, and Total categories yielded Instability Index values of 16%, 10%, and 15%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Instability Index values of 11%, 14%, and 12%, respectively.

When comparing fourth grade DIBELS category scores with fifth grade PSSA scores, fall administrations of ORF, Daze, and Total categories yielded Instability Index values of 10%, 17%, and 16%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Instability Index values of 13%, 14%, and 17%, respectively.

Table 8

Statistical Tests of Significance Comparing the Improvement Index Values Based on Fourth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Improvement based on Grade 4 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	15	15	-0.102	0.919
Daze	12	12	0.211	0.833
Total	8	12	-0.712	0.477
Fall	D1	D2		
ORF (D1) vs Daze (D2)	15	12	0.023	0.982
ORF (D1) vs Total (D2)	15	8	0.766	0.444
Daze (D1) vs Total (D2)	12	8	0.712	0.477
Spring	D1	D2		
ORF (D1) vs Daze (D2)	15	12	0.342	0.732
ORF (D1) vs Total (D2)	15	12	0.119	0.905
Daze (D1) vs Total (D2)	12	12	-0.211	0.833

Table 9

Instability Index Values Based on DIBELS and PSSA Category Scores

	Instability Index					
	Grade 4 DIBELS and Grade 4 PSSA		Grade 5 DIBELS and Grade 5 PSSA		Grade 4 DIBELS and Grade 5 PSSA	
DIBELS	Fall	Spring	Fall	Spring	Fall	Spring
ORF	13%	20%	16%	11%	10%	13%
Daze	13%	16%	10%	14%	17%	14%
Total	16%	15%	15%	12%	16%	17%

As shown in Table 10, Fisher's exact tests indicated that there were no significant differences in Instability Index values between fourth grade fall and spring administrations of ORF, Daze, and Total score categories. In addition, all comparisons of the Instability Index between ORF, Daze, and Total score categories within fall and spring of fourth grade were nonsignificant.

There were no significant differences in Instability Index values between fifth grade fall and spring administrations of ORF, Daze, and Total score categories. These results are displayed in Table 11. All comparisons of the Instability Index between ORF, Daze, and Total score categories within fall and spring of fifth grade were nonsignificant.

As shown in Table 12, when comparing fourth grade DIBELS category scores with fifth grade PSSA category scores, there were no significant differences between Instability Index values for the fall and spring administrations of DIBELS category scores. In addition, there were no significant differences between Instability Index values of DIBELS categories within fall and spring.

Table 10

Statistical Tests of Significance Comparing the Instability Index Values Based on Fourth Grade DIBELS Scores and Fourth Grade PSSA Scores

	Instability based on Grade 4 DIBELS and Grade 4 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	5	8	-0.860	0.389
Daze	6	7	-0.387	0.699
Total	8	7	0.106	0.916
Fall	D1	D2		
ORF (D1) vs Daze (D2)	5	6	-0.031	0.975
ORF (D1) vs Total (D2)	5	8	-0.421	0.674
Daze (D1) vs Total (D2)	6	8	0.410	0.682
Spring	D1	D2		
ORF (D1) vs Daze (D2)	8	7	0.489	0.625
ORF (D1) vs Total (D2)	8	7	0.583	0.560
Daze (D1) vs Total (D2)	7	7	0.091	0.928

Table 11

Statistical Tests of Significance Comparing the Instability Index Values Based on Fifth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Instability based on Grade 5 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	7	5	0.580	0.562
Daze	5	7	-0.489	0.625
Total	6	5	0.324	0.746
Fall	D1	D2		
ORF (D1) vs Daze (D2)	7	5	0.724	0.469
ORF (D1) vs Total (D2)	6	6	0.143	0.886
Daze (D1) vs Total (D2)	5	7	-0.579	0.563
Spring	D1	D2		
ORF (D1) vs Daze (D2)	5	7	-0.345	0.730
ORF (D1) vs Total (D2)	5	6	-0.142	0.887
Daze (D1) vs Total (D2)	6	5	0.239	0.811

Table 12

Statistical Tests of Significance Comparing the Instability Index Values Based on Fourth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Instability based on Grade 4 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	5	6	-0.280	0.780
Daze	8	6	0.491	0.623
Total	8	8	-0.183	0.855
Fall	D1	D2		
ORF (D1) vs Daze (D2)	5	9	-0.835	0.404
ORF (D1) vs Total (D2)	5	9	-0.666	0.505
Daze (D1) vs Total (D2)	8	8	0.183	0.855
Spring	D1	D2		
ORF (D1) vs Daze (D2)	5	6	-0.154	0.878
ORF (D1) vs Total (D2)	5	8	-0.632	0.527
Daze (D1) vs Total (D2)	6	8	-0.491	0.623

The proportion of students identified as Not Proficient with fall and spring ORF, Daze, and Total category scores and who earned PSSA scores also in the Not Proficient range is reported as the Sensitivity Index. Sensitivity Index values are displayed in Table 13.

Table 13

Sensitivity Index Values Based on DIBELS and PSSA Category Scores

	Sensitivity Index					
	Grade 4 DIBELS and Grade 4 PSSA		Grade 5 DIBELS and Grade 5 PSSA		Grade 4 DIBELS and Grade 5 PSSA	
DIBELS	Fall	Spring	Fall	Spring	Fall	Spring
ORF	81%	69%	76%	84%	84%	80%
Daze	77%	73%	84%	76%	68%	76%
Total	69%	73%	76%	80%	68%	68%

Fourth grade fall administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 81%, 77%, and 69%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 69%, 73%, and 73%, respectively.

Fifth grade fall administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 76%, 84%, and 76%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 84%, 76%, and 80%, respectively.

When comparing fourth grade DIBELS category scores with fifth grade PSSA scores, fall administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 84%, 68%, and 68%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Sensitivity Index values of 80%, 76%, and 68%, respectively.

Fisher's exact tests indicated that there were no significant differences in Sensitivity Index values between fourth grade fall and spring administrations of ORF, Daze, and Total score categories. These results are displayed in Table 14. In addition, all comparisons of the Sensitivity Index between ORF, Daze, and Total score categories within fall and spring of fourth grade were nonsignificant.

There were no significant differences in Sensitivity Index values between fifth grade fall and spring administrations of ORF, Daze, and Total score categories. These results are displayed in Table 15. All comparisons of the Sensitivity Index between ORF, Daze, and Total score categories within fall and spring of fifth grade were nonsignificant.

As shown in Table 16, when comparing fourth grade DIBELS category scores with fifth grade PSSA category scores, there were no significant differences between Sensitivity Index values for the fall and spring administrations of DIBELS category scores. In addition, there were no significant differences between Sensitivity Index values of DIBELS categories within fall and spring.

Table 14

Statistical Tests of Significance Comparing the Sensitivity Index Values Based on Fourth Grade DIBELS Scores and Fourth Grade PSSA Scores

	Sensitivity based on Grade 4 DIBELS and Grade 4 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	21	18	0.961	0.336
Daze	20	19	0.320	0.749
Total	18	19	-0.306	0.760
Fall	D1	D2		
ORF (D1) vs Daze (D2)	21	20	0.340	0.734
ORF (D1) vs Total (D2)	21	18	0.961	0.337
Daze (D1) vs Total (D2)	20	18	0.625	0.532
Spring	D1	D2		
ORF (D1) vs Daze (D2)	18	19	-0.306	0.760
ORF (D1) vs Total (D2)	18	19	-0.306	0.760
Daze (D1) vs Total (D2)	19	19	0.000	1.000

Table 15

Statistical Tests of Significance Comparing the Sensitivity Index Values Based on Fifth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Sensitivity based on Grade 5 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	19	21	-0.658	0.511
Daze	21	19	0.658	0.511
Total	19	20	-0.320	0.749
Fall	D1	D2		
ORF (D1) vs Daze (D2)	19	21	-0.658	0.511
ORF (D1) vs Total (D2)	19	19	0.000	1.000
Daze (D1) vs Total (D2)	21	19	0.658	0.511
Spring	D1	D2		
ORF (D1) vs Daze (D2)	21	19	0.658	0.511
ORF (D1) vs Total (D2)	21	20	0.340	0.734
Daze (D1) vs Total (D2)	19	20	-0.320	0.749

Table 16

Statistical Tests of Significance Comparing the Sensitivity Index Values Based on Fourth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Sensitivity based on Grade 4 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	21	20	0.340	0.734
Daze	17	19	-0.601	0.548
Total	17	17	0.000	1.000
Fall	D1	D2		
ORF (D1) vs Daze (D2)	21	17	1.251	0.211
ORF (D1) vs Total (D2)	21	17	1.251	0.211
Daze (D1) vs Total (D2)	17	17	0.000	1.000
Spring	D1	D2		
ORF (D1) vs Daze (D2)	20	19	0.320	0.749
ORF (D1) vs Total (D2)	20	17	0.918	0.359
Daze (D1) vs Total (D2)	19	17	0.601	0.548

The proportion of students identified as Proficient with fall and spring ORF, Daze, and Total category scores who earned PSSA scores also in the Proficient range is reported as the Specificity Index. These values are displayed in Table 17.

Fourth grade fall administrations of ORF, Daze, and Total categories yielded Specificity Index values of 69%, 82%, and 86%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Specificity Index values of 65%, 76%, and 80%, respectively.

Fifth grade fall administrations of ORF, Daze, and Total categories yielded Specificity Index values of 64%, 72%, and 70%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Specificity Index values of 66%, 74%, and 72%, respectively.

When comparing fourth grade DIBELS category scores with fifth grade PSSA scores, fall administrations of ORF, Daze, and Total categories yielded Specificity Index values of 70%, 76%, and 84%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Specificity Index values of 70%, 76%, and 76%, respectively.

Table 17

Specificity Index Values Based on DIBELS and PSSA Category Scores

	Specificity Index					
	Grade 4 DIBELS and Grade 4 PSSA		Grade 5 DIBELS and Grade 5 PSSA		Grade 4 DIBELS and Grade 5 PSSA	
DIBELS	Fall	Spring	Fall	Spring	Fall	Spring
ORF	69%	65%	64%	66%	70%	70%
Daze	82%	76%	72%	74%	76%	76%
Total	86%	80%	70%	72%	84%	76%

As shown in Table 18, Fisher's exact tests indicated that there were no significant differences in Specificity Index values between fourth grade fall and spring administrations of ORF, Daze, and Total score categories. All comparisons of the Specificity Index between ORF, Daze, and Total score categories within fall and spring of fourth grade were nonsignificant. However, the difference in Specificity Index for the fall administrations of ORF and Total category score was approaching significance (p = .053).

There were no significant differences in Specificity Index values between fifth grade fall and spring administrations of ORF, Daze, and Total score categories. These results are displayed in Table 19. In addition, all comparisons of the Specificity Index between ORF, Daze, and Total score categories within fall and spring of fifth grade were nonsignificant.

As shown in Table 20, when comparing fourth grade DIBELS category scores with fifth grade PSSA category scores, there were no significant differences between Specificity Index values for the fall and spring administrations of DIBELS category scores. In addition, there were no significant differences between Specificity Index values of DIBELS categories within fall and spring.

Table 18

Statistical Tests of Significance Comparing the Specificity Index Values Based on Fourth Grade DIBELS Scores and Fourth Grade PSSA Scores

	Specificity based on Grade 4 DIBELS and Grade 4 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	34	32	0.431	0.667
Daze	40	37	0.739	0.460
Total	42	39	0.800	0.424
Fall	D1	D2		
ORF (D1) vs Daze (D2)	34	40	-1.409	0.159
ORF (D1) vs Total (D2)	34	42	-1.937	0.053
Daze (D1) vs Total (D2)	40	42	-0.547	0.584
Spring	D1	D2		
ORF (D1) vs Daze (D2)	32	37	-1.107	0.268
ORF (D1) vs Total (D2)	32	39	-1.583	0.113
Daze (D1) vs Total (D2)	37	39	-0.484	0.628

Table 19

Statistical Tests of Significance Comparing the Specificity Index Values Based on Fifth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Specificity based on Grade 5 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	32	33	-0.210	0.834
Daze	36	37	-0.225	0.822
Total	35	36	-0.220	0.826
Fall	D1	D2		
ORF (D1) vs Daze (D2)	32	36	-0.857	0.391
ORF (D1) vs Total (D2)	32	35	-0.638	0.524
Daze (D1) vs Total (D2)	36	35	-0.638	0.524
Spring	D1	D2		
ORF (D1) vs Daze (D2)	33	37	-0.873	0.383
ORF (D1) vs Total (D2)	33	36	-0.649	0.516
Daze (D1) vs Total (D2)	37	36	0.225	0.822

Table 20

Statistical Tests of Significance Comparing the Specificity Index Values Based on Fourth Grade DIBELS Scores and Fifth Grade PSSA Scores

	Specificity based on Grade 4 DIBELS and Grade 5 PSSA (*n*)			
Fall vs. spring	Fall	Spring	Fisher's *z*	*p*
ORF	35	35	0.000	1.000
Daze	38	38	0.000	1.000
Total	42	38	1.000	0.317
Fall	D1	D2		
ORF (D1) vs Daze (D2)	35	38	-0.676	0.499
ORF (D1) vs Total (D2)	35	42	-1.663	0.096
Daze (D1) vs Total (D2)	38	42	-1.000	0.317
Spring	D1	D2		
ORF (D1) vs Daze (D2)	35	38	-0.676	0.499
ORF (D1) vs Total (D2)	35	38	-0.676	0.499
Daze (D1) vs Total (D2)	38	38	0.000	1.000

Kappa Index values indicate the percentage of increase over chance represented by the overall percentage of agreement between DIBELS categories and the PSSA. Kappa Index values are shown in Table 21.

Table 21

Kappa Index Values Based on DIBELS and PSSA Category Scores

	Kappa Index					
	Grade 4 DIBELS and Grade 4 PSSA		Grade 5 DIBELS and Grade 5 PSSA		Grade 4 DIBELS and Grade 5 PSSA	
DIBELS	Fall	Spring	Fall	Spring	Fall	Spring
ORF	46%	32%	36%	44%	49%	45%
Daze	57%	46%	51%	47%	42%	49%
Total	55%	51%	42%	48%	52%	42%

Fourth grade fall administrations of ORF, Daze, and Total categories yielded Kappa Index values of 46%, 57%, and 55%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Kappa Index values of 32%, 46%, and 51%, respectively.

Fifth grade fall administrations of ORF, Daze, and Total categories yielded Kappa Index values of 36%, 51%, and 42%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Kappa Index values of 44%, 47%, and 48%, respectively.

When comparing fourth grade DIBELS category scores with fifth grade PSSA scores, fall administrations of ORF, Daze, and Total categories yielded Kappa Index values of 49%, 42%, and 52%, respectively. Spring administrations of ORF, Daze, and Total categories yielded Kappa Index values of 45%, 49%, and 42%, respectively.

It was hypothesized that spring DIBELS scores would be more predictive of PSSA results than fall DIBELS scores. It was anticipated that fifth grade DIBELS scores would be more predictive of fifth grade PSSA scores than fourth grade DIBELS scores. The effects of instructional efforts and intervention should be more apparent as the time between administration of DIBELS and PSSA increases, and student performance should theoretically improve over time. In order to examine the overall relationship between DIBELS categories and PSSA performance, student performance data on fall and spring administrations of DIBELS and on PSSA were sorted into performance patterns. The results are displayed in Table 22.

Table 22

Student Performance Patterns

	Grade 4 DIBELS and Grade 4 PSSA			Grade 5 DIBELS and Grade 5 PSSA			Grade 4 DIBELS and Grade 5 PSSA		
	ORF	Daze	Total	ORF	Daze	Total	ORF	Daze	Total
Consistently Non-Proficient									
N – N – N	18	16	15	18	18	18	19	15	14
Negative Change									
P – N – N	0	3	4	3	1	2	1	4	3
P – P – N	5	3	4	3	3	4	3	3	5
N – P – N	3	4	3	1	3	1	2	2	3
Positive Change									
P – N – P	3	7	5	0	3	0	2	6	6
N – N – P	14	5	5	17	10	14	13	6	6
N – P – P	1	4	2	1	4	1	2	6	2
Consistently Proficient									
P – P – P	31	33	37	32	33	35	33	33	36

When comparing Grade 4 ORF with the PSSA, of the 32 students who did not obtain Proficient scores on either the fall or spring ORF administrations, 18 did not obtain a Proficient score on the PSSA. However, 14 of those students were able to obtain a Proficient score. Four students who obtained Not Proficient ORF scores in the fall obtained Proficient ORF scores in the spring. Among those four students, three obtained a Not Proficient score on the PSSA, and one obtained a Proficient score. Three students who obtained Proficient ORF scores in the fall obtained Not Proficient ORF scores in the spring and Proficient scores on the PSSA. A total of 36 students obtained Proficient ORF scores in the fall and spring. Of those students, 31 were able to achieve Proficient scores on the PSSA.

For Grade 4 Daze, among the 21 students who did not obtain Proficient scores on either fall or spring Daze administration, five were able to achieve Proficient PSSA scores. Of the 36 students who achieved Proficient scores on fall and spring administrations, 33 achieved Proficient PSSA scores. Among the eight students who obtained Not Proficient Daze scores in the fall and Proficient Daze scores in the spring, four achieved Proficient PSSA scores. Of the 10 students who obtained Proficient Daze scores in the fall and Not Proficient Daze scores in the spring, seven were able to achieve Proficient PSSA scores.

For Grade 4 Total category, among the 20 students who did not obtain Proficient scores on either fall or spring Total administration, five were able to achieve Proficient PSSA scores. Of the 41 students who achieved Proficient scores on fall and spring administrations, 37 achieved Proficient PSSA scores. Among the five students who obtained Not Proficient Total scores in the fall and Proficient Total scores in the spring, two achieved Proficient PSSA scores. Of the nine students who obtained Proficient Total scores in the fall and Not Proficient Total scores in the spring, five were able to achieve Proficient PSSA scores.

When comparing Grade 5 ORF with PSSA, among the 35 students who did not obtain Proficient scores on either fall or spring ORF administration, 17 were able to achieve Proficient PSSA scores. Of the 35 students who achieved Proficient scores on fall and spring administrations, 32 achieved Proficient PSSA scores. Of the two students who obtained Not Proficient ORF scores in the fall and Proficient ORF scores in the spring, one achieved a Proficient PSSA score. Of the three students who obtained Proficient ORF scores in the fall and Not Proficient ORF scores in the spring, none were able to achieve Proficient PSSA scores.

For Grade 5 Daze, among the 28 students who did not obtain Proficient scores on either fall or spring Daze administration, 10 were able to achieve Proficient PSSA scores. Of the 36 students who achieved Proficient scores on fall and spring administrations, 33 achieved Proficient PSSA scores. Among the seven students who obtained Not Proficient Daze scores in the fall and Proficient Daze scores in the spring, four achieved Proficient PSSA scores. Of the four students who obtained Proficient Daze scores in the fall and Not Proficient Daze scores in the spring, three were able to achieve Proficient PSSA scores.

For Grade 5 Total category, among the 32 students who did not obtain Proficient scores on either fall or spring Total administration, 14 were able to achieve Proficient PSSA scores. Of the 39 students who achieved Proficient scores on fall and spring administrations, 35 achieved Proficient PSSA scores. Of the two students who obtained Not Proficient Total scores in the fall and Proficient Total scores in the spring, one achieved a Proficient PSSA score. Of the two students who obtained Proficient Total scores in the fall and Not Proficient Total scores in the spring, none were able to achieve Proficient PSSA scores.

When comparing Grade 4 ORF with Grade 5 PSSA, among the 32 students who did not obtain Proficient scores on either fall or spring ORF administration, 13 were able to achieve Proficient PSSA scores. Of the 36 students who achieved Proficient scores on fall and spring administrations, 33 achieved Proficient PSSA scores. Among the four students who obtained Not Proficient ORF scores in the fall and Proficient ORF scores in the spring, two achieved a Proficient PSSA score. Of the three students who obtained Proficient ORF scores in the fall and Not Proficient ORF scores in the spring, two were able to achieve Proficient PSSA scores.

When comparing Grade 4 Daze with Grade 5 PSSA, among the 21 students who did not obtain Proficient scores on either fall or spring Daze administration, six were able to achieve Proficient PSSA scores. Of the 36 students who achieved Proficient scores on fall and spring administrations, 33 achieved Proficient PSSA scores. Among the eight students who obtained Not Proficient Daze scores in the fall and Proficient Daze scores in the spring, six achieved Proficient PSSA scores. Of the 10 students who obtained Proficient Daze scores in the fall and Not Proficient Daze scores in the spring, six were able to achieve Proficient PSSA scores.

When comparing Grade 4 Total category with Grade 5 PSSA, among the 20 students who did not obtain Proficient scores on either fall or spring Total administration, six were able to achieve Proficient PSSA scores. Of the 41 students who achieved Proficient scores on fall and spring administrations, 36 achieved Proficient PSSA scores. Among the five students who obtained Not Proficient Total scores in the fall and

Proficient Total scores in the spring, two achieved a Proficient PSSA score. Of the nine students who obtained Proficient Total scores in the fall and Not Proficient Total scores in the spring, six were able to achieve Proficient PSSA scores.

CHAPTER 5

DISCUSSION

The PSSA results of the current study's sample were compared to statewide results. For both the fourth and fifth grade samples, the proportion of students identified as Not Proficient on the PSSA in this study's sample did not significantly differ from the proportion of students identified as Not Proficient statewide. This indicates that this study's sample is representative of statewide fourth and fifth grade PSSA results.

In this study, the number of students identified as at risk based on components of the DIBELS was considered. For fourth grade, the number of students identified as at risk of earning a Not Proficient score on the PSSA was similar when comparing fall DIBELS ORF (48%), Daze (39%), and Total (33%) category scores and when comparing spring DIBELS ORF (47%), Daze (41%), and Total (39%) category scores. Similarly, for fifth grade, the number of students identified as at risk was not significantly different between fall DIBELS ORF (49%), Daze (47%), and Total (45%) category scores or between spring DIBELS ORF (51%), Daze (43%), and Total (45%) category scores. Results indicate that these components of the DIBELS identify similar percentages of students at risk. However, overall, ORF did identify the highest percentage of students as at risk. Within each grade, the number of students identified as at risk across fall and spring administrations was consistent.

Rate of improvement, or the proportion of students identified as at risk on the DIBELS who earned PSSA scores in the Proficient range, was also similar between the components of the DIBELS, as well as across fall and spring administrations for both grade levels and across grade levels. Results suggest that the ORF, Daze, and Total categories are similar indicators of improvement. However, ORF identifies slightly more students as at risk who actually earn PSSA scores in the Proficient range. This may suggest an over-identification of students at risk, compared to Daze and Total categories.

There were also no statistically significant differences in indication of instability of the DIBELS measures. All three DIBELS components identified similar proportions of students as not at risk with DIBELS scores who then earned PSSA scores in the Not Proficient range. Instability values were similar between DIBELS components and between fall and spring administrations, as well as when considering values across grade levels. The instability values were low for all comparisons, indicating that it is rare for students to pass a DIBELS assessment but do poorly on the PSSA. Results demonstrate that ORF, Daze, and Total categories are similar gauges of instability and that instability is low for these measures.

Sensitivity, or ability to determine the proportion of students who earned PSSA scores in the Not Proficient range and were correctly identified as at risk on the DIBELS, was also similar between the three DIBELS components. There were no statistically significant differences in sensitivity values between ORF, Daze, and Total categories between fall and spring administrations and when comparing fourth grade DIBELS with fifth grade PSSA results. When considering fourth grade DIBELS with fifth grade PSSA results, ORF yielded the highest Sensitivity Index values, compared to Daze and Total categories. This may suggest that, over time, ORF may be a slightly more accurate predictor of at risk students who obtain Not Proficient PSSA scores. Overall, the sensitivity values were high, indicating that a considerable proportion of students who did poorly on the PSSA were correctly identified as at risk on a DIBELS assessment prior to the PSSA.

There were also no statistically significant differences between measure of specificity of the DIBELS components between fall and spring administrations and across grade levels. ORF, Daze, and Total categories showed similar ability to correctly identify students as not at risk who earned PSSA scores in the Proficient range. Compared to ORF and Daze, the Total score yielded slightly higher Specificity Index values when comparing fourth grade DIBELS with fourth grade PSSA scores and when comparing fourth grade DIBELS with fifth grade PSSA scores. ORF yielded the lowest Specificity Index values overall. The difference between the Specificity Index for the fourth grade fall administrations of ORF and Total category score was approaching significance (p = .053). These results may suggest that Daze and Total categories accurately identify slightly more students as not at risk. Overall, specificity values were high, which suggests that a large percentage of students were correctly identified as not at risk.

When considering the Kappa Index, or percentage of increase over chance, ORF, Daze, and Total category scores had similar rates of agreement with the PSSA. The DIBELS components appear to be comparable in predicting PSSA outcomes in terms of improvement over chance. ORF yielded the lowest Kappa Index values when comparing fourth grade DIBELS with fourth grade PSSA scores and when comparing fifth grade DIBELS with fifth grade PSSA scores. This may suggest that the ORF has a slightly lower accuracy in prediction of improvement over chance, compared to Daze and Total categories, when considering scores within a grade level.

Results were also similar overall for ORF, Daze, and Total categories when considering student performance patterns. The majority of students who performed poorly on both fall and spring DIBELS administrations also obtained Not Proficient PSSA scores. This was true for approximately half of the measures. Based on fourth grade ORF compared to fourth and fifth grade PSSA scores, fifth grade ORF, and fifth grade Total category, almost half of the students who performed poorly on both DIBELS administrations were able to achieve Proficient PSSA scores. ORF may identify a slightly higher proportion of students at risk who are able to earn PSSA scores in the Proficient range.

Of the students who performed well on fall DIBELS administrations but did poorly on spring DIBELS administrations, more were able to achieve Proficient scores than Not Proficient scores on the PSSA, except when considering fifth grade ORF and Total categories. Of the students who were Not Proficient on fall DIBELS administrations but Proficient on spring DIBELS administrations, the number of students who were able to achieve Proficient PSSA scores compared to Not Proficient scores was similar.

As would be anticipated, the majority of students who did well on both administrations of the DIBELS also performed well on the PSSA. Few students who did well on both fall and spring administrations of the DIBELS obtained Not Proficient PSSA scores. The majority of students who did not obtain at least one Proficient score on either the fall or spring DIBELS administration were actually more likely to obtain Proficient PSSA scores. This was true for all measures, except for fifth grade ORF and fourth and fifth grade Total categories.

Implications of the Findings

This study focused on the relationship between multiple components of the DIBELS and the extent to which each DIBELS measure accurately identifies students at risk for failure on the PSSA. This study examined the prediction of PSSA scores within and across fall and spring of fourth and fifth grades. Research on the predictive efficiency of CBM is important, due to the need for school districts to demonstrate student reading progress through performance on statewide tests.

Overall, DIBELS ORF, Daze, and Total category scores were similar in identifying students at risk for failure on the PSSA, as well as in the Improvement, Instability, Sensitivity, Specificity, and Kappa indices. The hypothesis that ORF, Daze, and Total categories would not significantly differ in the identification of percentage of students at risk, as well as in the index values, was supported. There were no statistically significant differences between ORF, Daze, and Total category scores.

The hypothesis that spring DIBELS scores would be more predictive of PSSA results than fall DIBELS scores and that fifth grade DIBELS scores would be more predictive of fifth grade PSSA scores than fourth grade DIBELS scores was not supported. There were no statistically significant differences between DIBELS fall and spring administrations. Furthermore, there were no statistically significant differences when comparing fourth grade DIBELS with fifth grade PSSA scores. This could be due to the study's small sample size. Also, data spanning a longer period could yield different results.

The hypothesis that ORF, Daze, and Total categories would be predictive of PSSA results was supported. The majority of students were accurately identified as at risk on a DIBELS assessment. The DIBELS is a valuable tool for school districts to use to predict student performance. The measures had low instability, meaning that it was unusual for students to pass a DIBELS assessment but perform in the Not Proficient range on the PSSA. Sensitivity was high, indicating that a substantial percentage of students were correctly identified as at risk on a DIBELS assessment. Specificity values were also high, suggesting that a large percentage of students were correctly identified as not at risk. Those who did well on both fall and spring DIBELS assessments typically earned Proficient PSSA scores, and those who were at risk on fall and spring DIBELS assessments typically earned PSSA scores in the Not Proficient range. There was, however, a substantial number of students identified as at risk on both fall and spring DIBELS who were able to achieve Proficient PSSA scores. This could be due to the effects of instructional efforts and intervention or the over-identification of students at risk.

Results suggest a slight over-identification of students at risk, based on the ORF, as well as the possibility of more accurate prediction over time, but this should be interpreted with caution due to the study's small sample size. The over-identification may be beneficial for school districts not to miss students who would benefit from intervention. Compared to Daze and Total categories, ORF identified the largest number of students as at risk. Also, ORF identified more students to be at risk for failure who actually earned Proficient PSSA scores. Compared to ORF, Daze and Total categories correctly identified slightly more students as not at risk. ORF also yielded the lowest rate of agreement with the PSSA within a grade level. Over time, ORF correctly identified the largest proportion of at risk students who earned Not Proficient PSSA scores.

However, comparisons were not significant, and the study has a small sample size. Therefore, meaningful conclusions cannot be firmly established.

Results suggest that DIBELS Total category was not more predictive of PSSA results than ORF and Daze. These results are consistent with previous research (Dynamic Measurement Group, 2010; Ferchalk et al., 2012; Good et al., 2011; University of Oregon Center on Teaching and Learning, 2012, 2013). School districts may not find the need to administer all components of the DIBELS in order to obtain a Total score. Administering specific individual DIBELS assessments may be more time and cost effective than and as efficient as administering all DIBELS components.

Limitations

One limitation of this study is the examination of a small sample size over a relatively limited period. This study analyzed only 2 years of data of a single cohort. The participants were from a single district in one state. Results may differ in other districts or other states. The results of this study may not generalize to other grade levels, districts, or states. Also, results may differ over a longer period with multiple years of data.

Another limitation of this study is the examination of only one CBM, the DIBELS. There are many different CBMs that can be explored. Other CBMs may have different relationships with statewide tests.

This study was limited by a lack of consideration of additional factors relating to student performance. There are many aspects relating to student performance that were not measured in this study and that could have influenced results. Factors such as special education status, student motivation, and family support, as well as type of intervention, intervention intensity, and intervention duration may affect student academic functioning.

Another limitation is the use of categorical data. The measures use cut-off scores, and data were combined and collapsed into dichotomous categorical scores. This allowed for straightforward comparison of measures. However, categorical data does not reflect the continuous scores within each category, which does not allow for the examination of differences between student performances within each category.

Future Directions

Future research is necessary to explore the predictive efficiency of the DIBELS and other CBMs. Research on the predictability of a variety of CBMs over a longer period would be helpful. Analyzing long-term prediction would be helpful for school districts to know how to use CBM most efficiently and effectively. The use of screening and progress monitoring with CBM over time may allow school districts to increase the number of students who earn Proficient scores on statewide tests. This would allow time for adjustment of intervention. It would be helpful for school districts to know which CBMs and which CBM components are most cost and time effective and have the most predictive efficiency with statewide tests.

It would be beneficial for future research to also consider the myriad factors that may influence student performance. Also, the analysis of continuous data would allow for examination of nuances in student performance within the categories of each measure. It is important for educational professionals to be aware of the complex relationship between CBM and statewide testing and the implications that this information has for student reading performance.

REFERENCES

Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities, 41*(1), 67-84. doi:2048/10.1177/0022219407310838

Algozzine, B., Marr, M. B., Kavel, R. L., & Dugan, K. K. (2009). Using peer coaches to build oral reading fluency. *Journal of Education for Students Placed at Risk, 14,* 256-270. doi:10.1080/10824660903375735

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first grade Response to Intervention (RTI) in reading. *Exceptional Children, 81*(1), 11-27. doi:10.1177/0014402914532234

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38*(2), 266-283. Retrieved from http://www.nasponline.org/publications/spr/sprissues.aspx

Ardoin, S. P., Eckert, T. L., & Cole, C. A. S. (2008). Promoting generalization of reading: Acomparison of two fluency-based interventions for improving general education students' oral reading rate. *Journal of Behavioral Education, 17,* 237-252. doi:10.1007/s10864-009066-1

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral readingfluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16,* 55-70. doi:10.1007/s10864-006-9020-z

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L, Resetar, J. L., . . . Williams, K. L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33,* 218-233. Retrieved from https://www.nasponline.org/ resources-and-publications

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). *The research building blocks for teaching children to read: Put reading first: Kindergarten through grade 3* (3rd ed.). Retrieved from http://www.nifl.gov

Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina End of Grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.

Bashir, A. S., & Hook, P. E. (2009). Fluency: A key link between word identification and comprehension. *Language, Speech, and Hearing Services in Schools, 40,* 196-200. doi:10.1044/0161-1461

Beaver, J. & Carter, M. (2011). *The Developmental Reading Assessment – second edition PLUS.* San Antonio, TX: Pearson Education, Inc.

Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency programs: The Helping Early Literacy with Practice Strategies (HELPS) program and the Great

Leaps K-2 reading program. *School Psychology Review, 39,* 137-155. Retrieved from http://www.naspweb.org/publications/index.html

Begeny, J. C., & Martens, B. K. (2006). Assisting low-performing readers with a group-based reading fluency intervention. *School Psychology Review, 35,* 91-107. Retrieved from http://www.naspweb.org/publications/index.html

Berninger, V. W., Abbott, R. D., Vermeulen, K., & Fulton, C. M. (2006). Paths to reading comprehension in at-risk second-grade readers. *Journal of Learning Disabilities, 39*(4), 334-351. doi:10.1177/00222194060390040701

Block, C. C., Parris, S. R., Reed, K. L., Whiteley, C. S., & Cleveland, M. D. (2009). Instructional approaches that significantly increase reading comprehension. *Journal of Educational Psychology, 101*(2), 262-281. doi:2048/10.1037/a0014319

Brown, V. L., Wiederholt, L., & Hammill, D. D. (2009). *The Test of Reading Comprehension* (4th ed.). San Antonio, TX: Pearson Education, Inc.

Broxterman, K., & Whalen, A. J. (2013). *RTI team building: Effective collaboration and data-based decision making.* New York, NY: Guilford Press.

Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report No. 1). Tallahassee, FL: Florida State University, Florida Center for Reading Research.

Burns, M., K., Scholin, S. E., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment, 28*(2), 102-114. doi:2048/10.1177/0734282909342374

Bursuck, B., & Blanks, B. (2010). Evidence-based early reading practices within a Response to Intervention system. *Psychology in the Schools, 47*(5), 421-431. doi:10.1002/pits.20480

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing, 28,* 1407-1425. doi:10.1007/s11145-015-9576-x

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323. doi:10.1207/S15326977EA0704_04

Cummings, K. D., Biancarosa, G., Schaper, A., & Reed, D. K. (2014). Examiner error in curriculum-based measurement of oral reading. *Journal of School Psychology, 52*(4), 361-375. doi:10.1016/j.jsp.2014.05.007

Curriculum Associates. (2017). *Phonics for Reading*. Retrieved from https://www.curriculumassociates.com/products/detail.aspx?title=PhonicsReading#ordernow

Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia, 59*, 34-54. doi:10.1007/s11881-009-0022-0

Daly, E. J., & Kupzyk, S. (2012). An investigation of student-selected and parent-delivered reading interventions. *Journal of Behavioral Education, 21,* 295-314. doi:10.1007/s10864-012-9149-x

Daly, E. J., Chafouleas, S., & Skinner, C. H. (2005). *Interventions for reading problems:*

Designing and evaluating effective strategies. New York, NY: Guilford Press.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/001440298505200303

Desoff, A. (2007). DIBELS draws doers and doubters. *District Administration*, *43*(8), 38-43. Retrieved from: www.districtadministration.com

Dynamic Measurement Group. (2010). *DIBELS Next Benchmark Goals and Composite Score*. Retrieved from http://dibels.org

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167-188. doi:10.1207/s1532799xssr0902_4

Elbaum, A., Arguelles, M. E., Campbell, Y., & Saleh, M. B. (2004). Effects of a student-reads-aloud accommodation on the performance of students with and without learning disabilities on a test of reading comprehension. *Exceptionality, 12*(2), 71-87. Retrieved from http://www.leaonline.com.ezproxy.pcom.edu:2048

EPS School Specialty. (2018). *EPS Products*. Retrieved from http://eps.schoolspecialty.com/ products

Feifer, S. G. (2008). Integrating Response to Intervention (RTI) with neuropsychology: A scientific approach to reading. *Psychology in the Schools*, *45*(9), 812-825. Retrieved from http://dx.doi.org.ezproxy.pcom.edu:2048/10.1002/pits.20328

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18,* 52-65. Retrieved from http://dx.doi.org.ezproxy.pcom.edu:2048/10.1521/scpq.18.1.52.20876

Ferchalk, M. R. (2013). *Test nationally, benchmark locally: Using local DIBELS benchmarks to predict performance on the PSSA* (Doctoral dissertation). Retrieved from ProQuest, UMI Dissertations. (3589383)

Ferchalk, M. R., Cogan-Ferchalk, J. R., & Richardson, F. (2012). DIBELS Next and performance on high-stakes assessments. *ASPP InSight, 33*(1), 4-6.

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40*(1), 3-22. Retrieved from http://www.nasponline.org/publications/spr/spr401index.aspx

Ford, J. W., Missall, K. N., Hosp, J. L., & Kuhle. (2017). Examining oral passage reading rate across three curriculum-based measurement tools for predicting grade-level proficiency. *School Psychology Review, 46*(4), 363-378. doi:10.17105/SPR-2016-0014.V46-4

Fuchs, D., & Fuchs, L. D. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93-99. doi:10.1598/RRQ.41.1.4

Fuchs, D., & Fuchs, L. S. (2009). Responsiveness to intervention: Multilevel assessment and instruction as early intervention and disability identification. *Reading Teacher, 63,* 50-252. doi:10.1598/RT.63.3.10

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71,* 7-21. doi:10.1177/001440290407100101

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256. doi:10.1207/S1532799XSSR0503_3

Gardiner, S. (2001). Ten minutes a day for silent reading. *Educational Leadership*, *59*(2), 32-35. Retrieved from: https://eric.ed.gov/?id=EJ634676

Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools, 46,* 539-551. doi:10.1002/pits.20396

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. 679-700). Washington, DC: National Association of School Psychologists.

Good, H. R., & 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., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2011). *DIBELS Next Technical Adequacy Brief.* Eugene, OR: Dynamic Measurement Group. Retrieved from http://dibels.org/

Good, R. H., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288. doi:10.1207/S1532799XSSR0503_4

Good, R. H., Simmons, D., Kame'enui, E., 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.

Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review, 34*, 184-201. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=34&issue=2

Hale, J. B. (2006). Implementing IDEIA 2004 with a three-tier model that includes response to intervention and cognitive assessment methods. *School Psychology Forum: Research in Practice*, *1*, 16-27. Retrieved from https://www.nasponline.org/resources-and-publications

Harvey, S., & Goudvis, A. (2018). The Comprehension Toolkit. *Heinemann*. Retrieved from https://www.heinemann.com/comprehensiontoolkit/

Herbert, C. R. (2004). Running Records every day. *Instructor, 114*(1), 30-71. Retrieved from http://teacher.scholastic.com/products/instructor/index.htm

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34,* 372-386. Retrieved from https://www.nasponline.org/resources-and-publications

Hintze, J. M., Callahan, J. E., Matthews, W. J., Williams, S. A. S., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, *31*(4), 540-553. Retrieved from http://www.nasponline.org/publications/spr/sprissues.aspx

Hoffman, A. R., Jenkins, J. E., & Dunlap, S. K. (2009). Using DIBELS: A survey of purposes and practices. *Reading Psychology, 30*, 1-16. doi:10.1080/02702710802274820

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34,* 9-26. Retrieved from https://www.nasponline.org/resources-and-publications

Hosp, M. K., & MacConnell, K. L. (2008). Best practices in curriculum-based evaluation in early reading. In A. Thomas & J. Grimes (Eds.), *Best Practices V* (pp. 377-396). Bethesda, MD: National Association of School Psychologists.

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *Reading Teacher, 58,* 702-714. doi:10.1598/RT.58.81

Kaufman, A. S. & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement – 3rd edition*. San Antonio, TX: Pearson Education, Inc.

Kendeou, P., McMaster, K. L., & Christ, T. (2016). Reading comprehension: Core components and processes. *Policy Insights From the Behavioral and Brain Sciences, 3*(1), 62-69. doi:10.1177/2372732215624707

Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014) Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of School Psychology, 52*, 377-405. doi:10.1016/j.jsp.2014.06.002

Kim, Y. S. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly*, *50*(4), 459-481. doi:10.1002/rrq.107

Kim, J. S., Vanderwood, M. L., & Lee, C. Y. (2016). Predictive validity of curriculum-based measures for English learners at varying English proficiency levels. *Educational Assessment, 21*(1), 1-18. doi:10.1080/10627197.2015.1127750

Klingbeil, D. A., Van Norman, E. R., & Nelson, P. M. (2017). Precision of curriculum-based measurement reading data: Considerations for multiple-baseline designs. *Journal of Behavioral Education, 26*(4), 433-451. doi:10.1007/s10864-017-9282-7

Langdon, T. (2004). DIBELS: A teacher-friendly basic literacy accountability tool for the primary classroom. *Teaching Exceptional Children, 37*(2), 54-58. doi:10.1177/004005990403700207

Language Circle Enterprises. (2018). *Project Read*. Retrieved from https://www.projectread.com/

Lennon, L. (2017). *Reading interventions for struggling readers* [PowerPoint slides].

Lipka, O. (2017). Reading fluency from grade 2-6: A longitudinal examination. *Reading and Writing, 30*(6), 1361-1375. doi:10.1007/s11145-017-9729-1

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology, 47*, 315-335. doi:10.1016/j.jsp.2009.04.003

McCloskey, G. (2016). *Neuropsychology of reading disabilities* [PowerPoint slides].

McCloskey, G., & Perkins, L. A. (2013). *Essentials of executive functions assessment*. Hoboken, NJ: John Wiley & Sons.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties*. New York, NY: Routledge Press.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*(2), 193-203. Retrieved from https://www.nasponline.org/resources-and-publications

McMaster, K. L., van den Broek, P., Espin, C. A., Pinto, V., Janda, B., Lam, E., . . . van Boekel, M. (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology, 40*, 28-40. doi:10.1016/j.cedpsych.2014.04.001

Meisinger, E. B., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annuals of Dyslexia, 60,* 1-17. doi:10.1007/s11881-009-0031-z

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research and Practice, 24*(4), 186-195. doi:10.1111/j.1540-5826.2009.00292.x

Mesmer, E. M., & Mesmer, H. E. (2008). Response to Intervention (RTI): What teachers of reading need to know. *Reading Teacher, 62*, 280-290. doi:10.1598/RT.62.4.1

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly, 43,* 336-354. doi:10.1598/RRQ.43.4.2

Munger, K. A., LoFaro, S. A., Kawryga, E. A., Sovocool, E. A., & Medina, S. Y. (2014). Does the Dynamic Indicators of Basic Earl Literacy Skills Next assessment take a "simple view" of reading? *Educational Assessment, 19,* 204-228. doi:10.1080/10627197.2014.934609

Musti-Rao, S., Hawkins, R. O., & Barkley, E. A. (2009). Effects of repeated readings on the oral reading fluency of urban fourth-grade students: Implications for practice. *Preventing School Failure, 54,* 12-23. Retrieved from http://heldref.metapress.com.ezproxy.pcom.edu:2048/openurl.asp?genre=article&id=doi:10.3200/PSFL.54.1.12-23

Nation, K., & Angell, P. (2006). Learning to read and learning to comprehend. *London Review of Education, 4*(1), 77-87. doi:10.1080/13603110600574538

National Center for Learning Disabilities [NCLD]. (2018). *RTI action network*. Retrieved from https://www.ncld.org/

National Center on Response to Intervention. (2018). *The essential components of RTI.* Retrieved from https://www.rti4success.org/

National Institute of Child Health and Human Development [NICHHD]. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence based assessment of the scientific research literature on reading and its implications for* *reading instruction.* Reports of the subgroups, NIH Publication NO.00-4754. Washington, DC: U.S. Government Printing Office.

Nelson, J. M. (2008). Beyond correlational analysis of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A classification validity study. *School Psychology Quarterly, 23*(4), 542-552. Retrieved from http://dx.doi.org.ezproxy.pcom.edu:2048/ 10.1037/a0013245

Northwest Evaluation Association [NWEA]. (2016). *Linking the Pennsylvania PSSA Assessments to NWEA MAP Tests*. Portland, OR: Northwest Evaluation Association.

O'Connor, R. E., & Klingner, J. (2010). Poor responders in RTI. *Theory into Practice, 49*, 297-304. doi:10.1080/00405841.2010.510758

O'Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology, 102*(1), 1-19. doi:10.1037/a0017488

Pearson Education, Inc. (2018). *Comprehension Plus*. Retrieved from https://www.pearsonschool.com/index.cfm?locator=PSZu68&PMDbProgramID=256

Pennsylvania Department of Education [PDE]. (2016). *2016 Pennsylvania System of School Assessment Technical Report*. Retrieved from http://www.pde.state.pa.us.

Pennsylvania Department of Education [PDE]. (2017). PSSA state level data. Retrieved from https://www.education.pa.gov/data-and-statistics/PSSA/Pages/default.aspx

Pennsylvania Department of Education [PDE]. (2018). *Pennsylvania System of School Assessment: 2009-2010 assessment handbook*. Harrisburg, PA. Retrieved from www.education.state.pa.us.

Pennsylvania Department of Education Standards Aligned System. (2014). *PA core English Language Arts, grades PreK-5*. Retrieved from http://www.pdesas.org

Pennsylvania State Data Center. (2018). *Special education data report: LEA performance on state performance plan (SPP) targets, school year 2016-2017*. Retrieved from https://penndata.hbg.psu.edu/penndata/documents/BSEReports/Public%20Reporting/2016_2017/PDF_Documents/Speced_Data_Report_SD374_Final.pdf

Peterson, K. M. H., & Shinn, M. R. (2002). Severe discrepancy models: Which best explains school identification practices for learning disabilities? *School Psychology Review, 31*(4), 459-476. Retrieved from https://www.nasponline.org/resources-and-publications

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *Reading Teacher, 58,* 510-519. doi:10.1598/RT.58.6.2

Powell, D., Higgins, H. J., Aram, R., & Freed, A. (2009). Impact of No Child Left Behind on curriculum and instruction in rural schools. *Rural Educator, 31*(1), 19-28. Retrieved from https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=EJ876130

PRO-ED. (2018). *Reading Milestones*. Retrieved from https://www.proedinc.com/

Prosser, A. L. (2015). *Concurrent validity and diagnostic accuracy of curriculum-based assessment: Comparing the DIBELS Next to the CTOPP-2* (Doctoral dissertation). Retrieved from ProQuest, UMI Dissertations. (3706393)

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research and Practice*, *25*(30), 110-123. doi:10.1111/j.1540-5826.2010.00309.x

Rasinski, T., Homan, S., & Biggs, M. (2009). Teaching reading fluency to struggling readers: Method, materials, and evidence. *Reading and Writing Quarterly, 25,* 192-204. doi:10.1080/10573560802683622

Reutzel, D. R., & Hollingsworth, P. M. (1993). Effects of fluency training on second graders' reading comprehension. *Journal of Educational Research, 86*(6), 325-331. Retrieved from http://www.heldref.org/pubs/jer/about.html

Riedel, B. W., & Samuels, S. J. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students (with commentary). *Reading Research Quarterly, 42,* 546-567. doi:10.1598/RRQ.42.4.5

Ritchey, K. D., Palombo, K., Silverman, R. D., & Speece, D. L. (2017). Effects of an informational text reading comprehension intervention for fifth-grade students. *Learning Disability Quarterly*, *40*(2), 68-80. doi:10.1177/0731948716682689

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly, 20,* 304-317. doi:2048/10.1521/scpq. 2005.20.3.304

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343-366. doi:10.1016/j.jsp.2007.06.006

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*(3), 341-355. Retrieved from http://www.nasponline.org/publications/spr/sprissues.aspx

Russo, C., Tiegerman-Farber, E., & Radziewicz, C. (2009). *RTI guide: Making it work: strategies=solutions*. Port Chester, NY: Dude Pub.

Sattler, J. M. (2014). *Foundations of behavioral, social, and clinical assessment of children* (6th ed.). San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal, 107,* 429-448. doi:10.1086/518622

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24,* 19-35. doi:10.1177/0734282905285237

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

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.

Siah, P., & Kwok, W. (2010). The value of reading the effectiveness of sustained silent reading. *Clearing House, 83,* 168-174. doi:10.1080/00098650903505340

Simmons, D. C., Coyne, M. D., Kwok, O., McDonagh, S., Harn, B. A., & Kame, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*(2), 158-173. doi:10.1177/0022219407313587

Simmons, D. C., Kim, M., Kwok, O., Coyne, M. D., Simmons, L. E., Oslund, E.,…Rawlinson, D. (2015). Examining the effects of linking student performance and progression in a tier 2 kindergarten reading intervention. *Journal of Learning Disabilities, 48*(3), 255-270. doi:10.1177/0022219413497097

Stage, S., & Jacobsen, M. (2001). Predicting student success on a state-mandated performance based assessment using oral reading fluency. *School Psychology Review, 30,* 407-419. Retrieved from http://www.nasponline.org/publications/periodicals/spr/volume-30/volume-30-issue-3/predicting-student-success-on-a-state-mandated-performance-based-assessment-using-oral-reading-fluency

Swanson, H. L., & O'Connor, R. (2009). The role of working memory and fluency practice on reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities, 42,* 548-575. doi:10.1177/0022219409338742

Tarchi, C. (2015). Fostering reading comprehension of expository texts through the activation of readers' prior knowledge and inference-making skills. *International Journal of Educational Research, 72,* 80-88. doi:10.1016/j/ijer.2015.04.013

Tomlinson, C. A. (2011). *How to differentiate instruction in mixed-ability classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

Tran, L, Sanchez, T., Arellano, B., & Swanson, H. L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities, 44*(3), 283-295. doi:10.1177/0022219410378447

U.S. Department of Education. (2001). *No Child Left Behind Act of 2001, Pub. L. No. 107-110*. Retrieved from https://www2.ed.gov/policy/elsec/leg/esea02/index.html

U.S. Department of Education. (2004). *Individuals with Disabilities in Education Improvement Act*. Retrieved from https://sites.ed.gov/idea/

U.S. Department of Education. (2015). *Every Student Succeeds Act (ESSA)*. Retrieved from https://www.ed.gov/esea

University of Oregon Center on Teaching and Learning. (2009). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS).* Retrieved from https://dibels.uoregon.edu

University of Oregon Center on Teaching and Learning. (2012). *2012-2013 DIBELS data system update part 1: DIBELS Next composite score (Technical Brief No. 1202).* Retrieved from https://dibels.uoregon.edu/research#dibels

University of Oregon Center on Teaching and Learning. (2013). *Guide to unbundling the composite score and using the recommended goals*. Retrieved from https://dibels.uoregon.edu

University of Oregon Center on Teaching and Learning. (2018). *DIBELS Next.* Retrieved from https://dibels.uoregon.edu/assessment/dibels/dibels-next

Utchell, L. A., Schmitt, A. J., McCallum, E., McGoey, K. E., & Piselli, K. (2016). Ability of early literacy measures to predict future state assessment performance. (2016).

Journal of Psychoeducational Assessment, 34(6), 511-523. doi:10.1177/0734282915621221

van den Bosch, R. M., Espin, C. A., Chung, S., & Saab, N. (2017). Data-based decision-making: Teachers' comprehension of curriculum-based measurement progress-monitoring graphs. *Learning Disabilities Research & Practice, 32*(1), 46-60. doi:10.1111/ldrp.12122

VanDerHeyden, A. M., & Burns, M. K. (2010). *Essentials of response to intervention*. Hoboken, NJ: Wiley.

Van Norman, E. R., & Christ, T. J. (2016). How accurate are interpretations of curriculum-based measurement progress monitoring data? Visual analysis versus decision rules. *Journal of School Psychology, 58*, 41-55. doi:10.1016/j/jsp.2016.07.003

Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2015). What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading, 38*(3), 213-225. doi:10.1111/1467-9817.12024

Voyager Sopris Learning. (2018). *Six-Minute Solution*. Retrieved from http://store.voyagersopris.com/six-minute-solution/

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541-561. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=36&issue=4

Wanzek, J., & Vaughn, S. (2011). Is a three-tier reading intervention model associated with reduced placement in special education? *Remedial and Special Education, 32*(2), 167-175. Retrieved from http://dx.doi.org/10.1177/0741932510361267

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades k-3. *Educational Psychology Review, 29,* 551-576. Retrieved from https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED579894

Wechsler, D. (2009). *The Wechsler Individual Achievement Test – third edition*. San Antonio, TX: Pearson Education, Inc.

Whitten, E., Esteves, K. J., & Woodrow, A. (2009). *RTI success: Proven tools and strategies for schools and classrooms*. Minneapolis, MN: Free Spirit Pub.

Widerholt, J., & Bryant, B. (2012). *Gray Oral Reading Test – fifth edition*. Austin, TX: PRO-ED.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207-214. doi:10.1177/07419325050260040301

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. San Antonio, TX: Pearson Education, Inc.

Winsor Learning. (2017). *Sonday System.* Retrieved from https://www.winsorlearning.com/

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovette, M. W., Wolf, M., . . . Schwanenflugel, P. (2010). The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. *Language, Speech, and Hearing Services in Schools, 41,* 340-348. Retrieved from http://dx.doi.org/10.1044/0161-1461 (2009/08-0093)

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment*, *11,* 85-104. doi:2048/10.1207/ s15326977ea1102_1

Woodcock, R., Mather, N., & McGrew, K. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing, Inc.

Woodcock, R., Mather, N., & Schrank, F. A. (2005). *Woodcock Diagnostic Reading Battery*. Rolling Meadows, IL: Riverside Publishing, Inc.

Woodward, M. M., & Talbert-Johnson, C. (2009). Reading intervention models: Challenges of classroom support and separated instruction. *Reading Teacher, 63*(3), 190-200. doi:10.1598/RT.63.3.2

Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties.* New York, NY: Springer.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24,* 4-12. doi:10.1111/j.1745-3992.2005.00014.x

Yurick, A. L., Robinson, P. D., Carledge, G., Lo, Y., & Evans, T. L. (2006). Using peer mediated repeated readings as a fluency-building activity for urban learners. *Education and Treatment of Children, 29,* 469-506. Retrieved from http://www.educationandtreatmentofchildren.net/contents/index.html

Zabrucky, K. M., Moore, D., Agler, L. M. L., & Cummings, A. M. (2015). Students' metacomprehension knowledge: Components that predict comprehension

performance. *Reading Psychology*, *36*(7), 627-642.
doi:10.1080/02702711.2014.950536