

ESTABLISHING PREDICTIVE VALIDITY FOR ORAL PASSAGE READING
FLUENCY AND VOCABULARY CURRICULUM-BASED
MEASURES (CBMS) FOR SIXTH GRADE STUDENTS

by

BRIAN R. MEGERT

A DISSERTATION

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

June 2010

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Brian Megert

Title:

"Establishing Predictive Validity for Oral Passage Reading Fluency and Vocabulary Curriculum-Based Measures (CBMs) for Sixth Grade Students "

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

Gerald Tindal, Chairperson, Educational Methodology, Policy, and Leadership
Paul Yovanoff, Member, Educational Methodology, Policy, and Leadership
Keith Hollenbeck, Member, Educational Methodology, Policy, and Leadership
Jean Stockard, Outside Member, Planning Public Policy & Mgmt

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

June 14, 2010

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

© 2010 Brian R. Megert

An Abstract of the Dissertation of

Brian R. Megert for the degree of Doctor of Education

in the Department of Educational Methodology, Policy, and Leadership

to be taken June 2010

Title: ESTABLISHING PREDICTIVE VALIDITY FOR ORAL PASSAGE READING
FLUENCY AND VOCABULARY CURRICULUM-BASED MEASURES
(CBMS) FOR SIXTH GRADE STUDENTS

Approved: _____
Dr. Gerald Tindal

In recent years, state and national policy created the need for higher accountability standards for student academic performance. This increased accountability creates an imperative to have a formative assessment system reflecting validity in inferences about the effectiveness of instruction and performance on statewide large-scale assessments. Curriculum-based measurement (CBM) satisfies both functions. However, research shows the predictive power of oral passage reading fluency (PRF) diminishes in middle and high school. Because of the decreased predictive validity of PRF in the upper grade levels, additional reading CBMs should be explored. This study compares PRF and Vocabulary CBM data for all sixth grade

students in a school district using two statistical procedures: correlation and regression. The correlation coefficients were moderately high among PRF, Vocabulary CBM, and the Reading test in Oregon Assessment of Knowledge and Skills (OAKS). A regression analysis indicated that the Vocabulary CBM explained more variance than PRF in predicting reading performance on OAKS. A second multiple regression analysis introduced three non-performance indicators (Gender, Attendance, and NCLB At-Risk), along with the two CBMs (Vocabulary and PRF). The second regression results revealed that Vocabulary again was more predictive than PRF, Gender, Attendance, or NCLB At-Risk. At-Risk status was the only non-performance indicator that was significant. All the findings have been discussed within the context of understanding reading skills using CBMs and their relation to performance on a large-scale test used for accountability. The findings have been framed as part of an information system that allows schools and districts to better tailor staffing, instruction, and schedules to student needs. Suggestions for future research also have been discussed, particularly in enhancing the predictions on large-scale test outcomes using a variety of CBMs.

CURRICULUM VITAE

NAME OF AUTHOR: Brian Richard Megert

PLACE OF BIRTH: Portland, Oregon

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon

DEGREES AWARDED:

Doctor of Education, Educational Methodology, Policy, and Leadership, 2010,
University of Oregon

Master of Science, Special Education, 1999, University of Oregon

Bachelor of Science, Psychology (Biology Minor), 1997, University of Oregon

AREAS OF SPECIAL INTEREST:

Special Education as Part of Building-Wide Efforts to Improve Student Achievement
Response to Intervention (RtI) at the Elementary and Middle School Level
Train and Support Teachers in the Area of Reading RtI
Reading Instruction for Elementary and Middle School Students
Completed *Trainer-of-Trainer Workshop Series* for REWARDS Curriculum
Completed *Institute on Becoming an Effective DI Trainer*, Summer, 2006
Services for Adult Students with Disabilities

PROFESSIONAL EXPERIENCE:

Special Programs Facilitator, Springfield School District, Springfield, Oregon, 2006-
Present

Coordinated Services and Programs for Students Throughout the District
Kindergarten through age 21

Elementary Special Education Teacher, Thurston Elementary, Springfield School
District, Springfield, Oregon, 2002-2006

Classroom for students with Emotional and Behavioral Concerns

Middle School Special Education Teacher, Clear Creek Middle School, Gresham-
Barlow School District, Gresham Oregon, 2001-2002
Resource Teacher for Students with a Wide-Range of Disabilities

Special Education Teacher for Grades 5-8, Meadow View School, Bethel School
District, Eugene, Oregon, 1999-2001
Resource Teacher for Students with a Wide-Range of Disabilities

ACKNOWLEDGEMENTS

First, I would like to thank my parents for instilling in me a drive for accomplishment and success. It means everything to me.

I would also like to thank all of the incredible professionals with whom I had the opportunity to work. The coursework during this program helped prepare me for further work in the field of K-12 education. Specifically, I am forever indebted to Dr. David Collins, Dr. Nancy Heapes, Dr. Matt Coleman, and Dr. Kathleen Scalise for imparting knowledge in a way that I had never experienced before. Also, Dr. Edward Kame'enui for demanding production and excellence that was an exceptional driving force in completing this project.

Additionally, each member of my committee helped improve this work. First, Dr. Gerald Tindal agreed to be my advisor during his sabbatical. He has taught me an incredible amount over the past decade and provided much needed feedback in all stages of the writing process. Also, Dr. Keith Hollenbeck took his valuable time to impart a great deal of technical advice, edits, and suggestions for improvements. Dr. Paul Yovanoff taught me invaluable information regarding my statistical analyses; he is a fantastic teacher and I would not have been able to decipher the results of this study without him. Last, I would like to thank my outside member, Dr. Jean Stockard. She presented me with vital knowledge and a unique perspective that would not have arose otherwise.

I dedicate this work to my absolutely incredible wife and daughter. Raina is the love of my life and the reason for my achievements, both personally and professionally. Sofia is the most amazing daughter a dad could hope to have and will accomplish anything and everything she chooses.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Policy Influence	3
Accountability Through Large-Scale Assessments	4
Validity and Use of Large-Scale Assessments: Purpose Matters	5
Inadequacy of Large-Scale Assessment Systems to Inform Instruction	7
The Utility of Reading CBMs	9
ORF as a Predictor of Future Performance	9
Vocabulary CBMs	10
Purpose of This Study	11
II. LITERATURE REVIEW	12
CBM to Inform Instruction and as a Predictor	13
CBM to Inform Instruction	13
Reading CBM as a Predictor of Large-Scale Outcome Measures	19
Vocabulary Acquisition, Assessment, and CBM	24
Vocabulary Acquisition and Assessment	24
Vocabulary CBM in Content Areas	28
Vocabulary as an Indicator of Generalized Reading Skills	30
Theory of Vocabulary as a Measurement of Generalized Reading Skills	31
Empirical Basis for Vocabulary as a Measure of Generalized Reading Skills	33
III. METHODOLOGY	35
Setting and Participants	35
Procedures	40
Training Procedures	40
Assessment Administration Procedures	41
Oregon Statewide Assessment	43
Measures	44
Passage Reading Fluency	44
Passage Reading Fluency Administration	45
Vocabulary Measure	46
Vocabulary Administration	47
Reading Portion of the Oregon Assessment of Knowledge and Skills	48

Chapter	Page
Analyses	49
IV. RESULTS	51
Cases Included and General Description	51
Analyzing for Multicollinearity.....	52
Research Question 1: Connection Among Measurement Variables	55
Question 2A: Predictive Nature of Performance Indicators	56
Question 3B: Predictive Nature With Additional Variables.....	58
Summary	61
V. DISCUSSION	62
Summary of Results From Analyses	62
The Connection Between Reading CBMs and OAKS	63
The Relative Predictive Nature of Two CBMs to OAKS.....	63
Limitations	64
Mortality	65
Grade Level	65
Standardization and Motivation.....	66
Measure of Instruction	67
Findings.....	68
PRF as a Predictor.....	68
Vocabulary as a Predictor	70
Practical Implications.....	71
Future Research.....	74
Conclusion.....	78
APPENDICES	
A. DISTRIBUTION OF SCORES FOR SIXTH GRADE OAKS, PRF, AND VOCABULARY.....	80
B. FALL SIXTH GRADE BENCHMARK PASSAGE, STUDENT AND ASSESSOR COPIES	83
C. FALL SIXTH GRADE BENCHMARK EXAMPLE VOCABULARY ITEMS.....	85
REFERENCES	86

LIST OF FIGURES

Figure	Page
1. Distribution of Reading OAKS Scores with Normal Curve Reference	80
2. Distribution of PRF Scores with Normal Curve Reference.....	81
3. Distribution of Vocabulary Scores with Normal Curve Reference.....	82

LIST OF TABLES

Table	Page
1. Percent Gender and Race/Ethnicity for Sixth, Seventh, and Eighth Grades	37
2. Percent Gender and Race/Ethnicity in Sixth Grade and District-Wide	38
3. Students With Disabilities Within Race/Ethnicity Subgroup.....	39
4. Descriptive Statistics of Assessment Results	52
5. Zero Order Correlation Matrix.....	53
6. Tolerance / VIF Matrix.....	55
7. ANOVA Statistics for Model 1, Using PRF and Vocabulary	56
8. Regression of OAKS-Reading on CBMs	57
9. Semi-Partial Correlations: OAKS-Reading on CBMs	57
10. Regression of OAKS-Reading on CBMs and Other Factors.....	59
11. Semi-Partial Correlations: OAKS-Reading on CBMs and Other Factors.....	60
12. Variance Accounted for Through Each Model.....	61

CHAPTER I

INTRODUCTION

Accurately defining and measuring reading skills has become increasingly critical in recent years (Baker et al., 2008). The requirements for high student achievement and the potential ramifications for low student achievement have increased dramatically in the past two decades (Herman, 2007). The summative high stakes of academic achievement, specifically reading achievement, has created the impetus for expanding and refining the use of formative reading curriculum-based measures (CBMs) (Deno, 2003). CBMs are used to accurately and efficiently measure reading skills and provide information about potential student performance on outcome measures, specifically statewide assessments (Good, Simmons, & Kame'enui, 2001). Because of the importance of the summative statewide assessments, districts, schools, and teachers must have the opportunity to adjust instruction to increase the chance of students meeting required benchmarks. Without measurement systems that provide these types of information, schools and districts will not have the ability to effectively serve the diverse needs of students that enter the education system.

Education agencies at the federal, state, and local level must have the opportunity to measure student learning in meaningful ways. Districts and schools are held accountable for student learning through summative, large-scale, standardized assessments (No Child Left Behind Act of 2001 [NCLB], 2002). Large-scale

assessments are constructed to measure whether students met state-defined information and developed the necessary academic skills to continue to progress through the educational system. Federal and state policies grant power to state departments of education to implement consequences to districts that do not demonstrate that their students are showing adequate growth (NCLB, 2002). Unfortunately, the large-scale assessments by which educators are held accountable for student learning do not provide adequate and timely information that allow teachers to adjust instruction to enhance student learning (Tindal, 2002). For this reason, CBMs are an important tool for educators. CBMs have the potential to provide reliable, valid, and timely information to teachers about student performance so instruction can be adjusted before students take summative outcome measures (Deno, 2003). CBMs can be a useful tool for enhancing the instruction for all students. But the utility of the results from CBMs can only be functional if the results are indicators of student learning in the larger constructs (Kame'enui et al., 2006). Recent research has demonstrated that oral reading fluency (ORF) is a reliable measure of reading skills and a valid predictor of future performance on large-scale outcome assessments for students in early elementary school.

In the following sections I will first supply information regarding the influence that state and federal policy has in the area of assessment. Second, I will discuss accountability through large-scale assessments, looking at both the validity and shortcomings of statewide assessments. Third, I will introduce the utility of reading CBMs in the context of accountability. Fourth, I will expand on the relevant research

pertaining to studies in the area of CBMs and how the utility of ORF as a predictive tool in the middle school grades diminishes. Finally, I will provide information regarding recent research in the area of CBMs that can supplement the use of fluency measures, specifically in the area of vocabulary. Ultimately, this gap in research will lead to my research questions regarding the predictive nature of fluency and vocabulary reading CBMs in relation to the reading portion of the Oregon statewide assessment for sixth grade students.

Policy Influence

In the 1980s, the nation's perspective about the education system shifted from confidence to skepticism when *A Nation At Risk* was published (National Commission on Excellence in Education, 1983). The report, fueled by a national concern of economic stability and security, declared that the U.S. education system failed our students and produced outcomes that were inadequate for students to succeed in college and the workforce. Current school reform found its genesis from this policy (Fowler, 2009). In Oregon, the state legislature passed the Oregon Education Act for the 21st Century in 1991 (Oregon School Boards Association [OSBA], 2005), which contributed to the establishment of the current summative assessment and accountability system in the state (Conley, 2007). Most states went through a similar process as pressures mounted on local and state governments to ensure high student achievement. As a continuation of the state and national educational reform effort, the No Child Left Behind (NCLB) Act of 2001 (NCLB, 2002) placed tremendous importance on assessment and accountability in the schools.

NCLB's focus on accountability was meant to increase student achievement. One way that the accountability increased was through scrutiny by public reporting of disaggregated district data for all subgroups (Oregon Revised Statute [ORS] 329.085). The subgroups include (a) students with disabilities, (b) English language learners, (c) students who are racial or ethnic minorities, and (d) students who are economically disadvantaged. In addition, states were provided the authority and obligation to establish sanctions for schools and districts that did not meet predetermined benchmarks of student achievement (ORS 329.105). Furthermore, states were required to adopt technically adequate student assessment systems (ORS 329.488). Student achievement benchmarks were nearly exclusively measured by student performance on large-scale, statewide assessments (Conley, 2007). This focus drastically increased the stakes of statewide assessments. While some may debate whether large-scale assessments measure student achievement in a meaningful way (Herman, 2007; Linn, 2002), it is clear that schools and districts are held accountable based on the results of the large-scale, statewide assessments (Tindal, 2002).

Accountability Through Large-Scale Assessments

At the policy level, accountability for school districts was traditionally designed to ensure that programs were in place. In the case of large, federal programs like the Individuals with Disabilities Education Act (IDEA) and Title grant programs, audits focused on whether monies were spent appropriately and resources were properly allocated (Fowler, 2009). Fowler observed that the NCLB legislation moved the focus from *equality of opportunity* to *equality of achievement*, which means accountability

measures shifted from program implementation to student achievement. In order to track student achievement at the school, district, and state levels, statewide assessment scores are disaggregated by subgroups and analyzed. Adequate performance by students in each of the subgroup categories is necessary, including students with (a) disabilities, (b) limited English proficiency (LEP), (c) underrepresented minority status, and (d) low socioeconomic status. States have the authority to sanction and ultimately reconstitute schools for continued lack of achievement overall or in any of the subgroup populations (Oregon Department of Education [ODE], 2009). Because the stakes have increased, the appropriate use of assessments has become increasingly important. Many statewide assessments have attributes that contribute to their valid use, but the assessments are only valid when the results are used and interpreted appropriately. When the results of large-scale assessments are used in ways for which they were not designed, the inadequacies of the assessments surface (Kane, 2002; Linn, 2002). When these inadequacies are revealed, the benefits of reading CBMs become much clearer.

Validity and Use of Large-Scale Assessments: Purpose Matters

Large-scale assessments, like the Oregon statewide assessment, are technically adequate when the results are used in specific ways (Kane, 2002). The Oregon Department of Education (ODE) provided in-depth information regarding the technical adequacy of the statewide assessment system. The report (ODE, 2007) included information regarding the development of the assessment, the standards on which the assessment was based, reliability, and several aspects of validity. The reliability information included the errors of the measurement at different score ranges. The report

also addressed concurrent validity by determining how well the results of the assessment correlated to nationally normed assessments, such as the California Achievement Test and the Iowa Test of Basic Skills. According to the technical document, the Oregon Assessment of Knowledge and Skills (OAKS) is adequate for providing information that can (a) provide instructionally useful evaluation of individual student progress toward mastery of the Academic Content Standards, (b) guide instructional program improvement, (c) ensure that the state is progressing toward the state and federal goals for high standards for all, and (d) inform the public.

Linn (2002) asserted that the most common use of the results of statewide assessments is measuring how well students progress toward state content standards. Linn also contended that results may be used to provide diagnostic information about student progress or for making high-stakes decisions about students. Tindal (2002) also argued that individual student progress toward content standards is the most valid use of large-scale assessments. Tindal stated further that because statewide assessment results are returned too late to inform any level of instruction, these scores are most useful for determining how much learning has taken place. He asserted that these large-scale assessment data are not useful for informing instruction.

Validity cannot be determined for an assessment alone, the valid use of the results or valid inferences can be made using the results. In other words, the interpretation determined by the proposed use of the assessment can be validated (Kane, 2002; Messick, 1995); the validity of an assessment is determined by the proposed use and the supporting argument. For example, the National Assessment of Educational

Progress (NAEP), a national large-scale assessment system, has been used to measure the relative success of state education systems for many years. It is considered to be a valid measurement tool at this level. It cannot, however, be used as an assessment for high-stakes decision-making at the individual student, district, or state level because the assessment uses results from a sample of students rather than assessing an entire population. In order for these assessments to cover a larger scope of uses, a different sampling methodology would be required.

Inadequacy of Large-Scale Assessment Systems to Inform Instruction

Large-scale assessments are present in all 50 states (NCLB, 2002). Tindal (2002) stated that there are trends nationally and internationally to increase the *amount and rigor* of large-scale assessments. Policy-makers and the public have supported this movement (Tindal). Despite this rigor, the results of large-scale assessments tend to hold limited utility in the context of classroom-based instruction and interventions (Fuchs & Fuchs, 1999). The assessment results are of limited utility because the large-scale assessments are administered yearly and are not delivered to teachers until after the majority of instruction has already occurred. Even if results were provided immediately, interpreting the large-scale assessments for instructional purposes would be suspect (Linn, 2002). Instead of using the results of summative assessments inappropriately, teachers and building administrators should use the results of formative assessments to inform instruction.

Assessment components necessary for informing instruction. Inexpensive formative measures hold instructional relevance and must be used to guide individual

student and classroom instruction and intervention strategies (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Good et al. (2001) explained that “for the purpose of informing instruction in time-efficient, instructionally relevant ways capable of altering students’ rates and levels of learning on critical indicators of reading, commercial standardized measures are severely limited, if not inappropriate” (p. 259). In addition to the inadequacy of large-scale assessment to inform instruction, Elmore (2004) theorized that the use of large-scale assessment scores alone for accountability is inadequate because the use of a *point in time* assessment score can distort conclusions about student knowledge. Elmore contended that scores on a test are simply a sample of knowledge, not necessarily knowledge of a domain. In these cases, instructional adjustments based on the single score may lead to counterproductive changes. Using formative assessment to inform instruction is likely more useful.

An alternative to large-scale assessment. Extensive research on CBMs through the 1980s (e.g., Deno, 1985; Marston, Fuchs, & Deno, 1986) established early reading skill measures as useful tools for informing instruction. The research community soon recognized fluency assessments as a valid measures of reading skill acquisition. As research continued in the area of fluency and automaticity, national level decision-makers endorsed and expanded the use of fluency measures for early reading (Adams, 1990). The National Reading Panel (NRP) report (Commission on Reading, 1985), NCLB (2002), and IDEA (2004) viewed fluency as one of the five essential areas of reading instruction and assessment. The endorsement at the national level reinforced the

findings from extensive research and validated the proposed uses of oral reading fluency as an accurate indicator of overall reading skill.

The Utility of Reading CBMs

As stated above, schools are measured by student performance on large-scale outcome assessments and, yet, those assessments are not useful for adjusting instructional practices in the classroom (Tindal, 2002). This being the case, formative assessments, specifically reading CBMs, must document information about the progression of overall reading skills and future performance on statewide assessments. Several studies have demonstrated that ORF is an indicator of overall reading skills in the elementary grades (e.g., Fuchs et al., 2001; Good et al., 2001). However, this utility tends to diminish in the upper grades (Fuchs et al., 2001), so further exploration of valid reading CBMs for middle school students, such as vocabulary, is necessary.

ORF as a Predictor of Future Performance

ORF has become an established and accurate tool for measuring reading skills at the early grades (Adams, 1990). In addition, the utility of ORF has been expanded as an accurate predictor of future success on large-scale, outcome assessments for elementary students (Good et al., 2001). Research has demonstrated that ORF is closely associated with comprehension (Burns et al., 2002) and that a large portion of students who meet particular thresholds on ORF measures tend to meet benchmark scores on grade-level, statewide assessments (Baker et al., 2008). Particularly in kindergarten through third grade, ORF was a very useful tool when measuring reading skill growth and predicting success on large-scale assessments.

As students progress through the grades and reading skills become more complex, the utility of ORF as a measure of overall reading skills and a predictor tended to diminish (Fuchs et al., 2001). Overall, there was a plateau effect for the skill of reading aloud fluently (Hasbrouck & Tindal, 2006). In addition, the correlation between ORF scores and performance on statewide assessments decreased for students in the middle school grades (Silbergitt, Burns, Madyun, & Lail, 2006). Because of the diminished utility of ORF as a CBM in the later grades, fluency measures became insufficient for understanding older students' reading skills. Additional information was required for teachers to make valid instructional decisions and determine whether students were likely to pass outcome measures. A complete formative assessment system should include alternative CBMs, in addition to reading fluency measures, for middle school students. One such alternative CBM could assess vocabulary.

Vocabulary CBMs

Vocabulary skills and knowledge are important aspects of reading development (Pearson, Hiebert, & Kamil, 2007). Reading is an intricate process that includes fluently decoding, but also involves interpreting and understanding text (Adams, 1990). Vocabulary is one of the necessary skills required for comprehension (Nagy & Anderson, 1984). While measuring vocabulary skills and vocabulary acquisition are complicated, several studies have demonstrated how vocabulary CBMs were correlated to student performance on outcome measures (Espin & Deno, 1994-95; Espin, Shin, & Busch, 2005; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Because the Espin studies primarily focused on content area assessments, they demonstrated that vocabulary

CBMs could provide information about future student performance on outcome measures. In addition, current theory and research suggested that vocabulary was an important component of developing more complex reading skills, and yet Pearson et al. (2007) stated that research has largely ignored assessments that can accurately measure vocabulary skills.

Purpose of This Study

This study was designed to address the gap in research in two areas. First, oral passage reading fluency in middle school grades has not been studied as thoroughly as in the elementary grades. Part of the analysis addressed how fluency CBMs administered to sixth grade students in the fall correlate to performance on the statewide assessment. Second, vocabulary CBMs were not commonly used as a benchmarking or progress-monitoring tool. This study addressed the relationship among vocabulary measures, fluency CBMs, and the statewide assessment. In addition, I established the relative predictive nature of the fluency and vocabulary CBMs in relation to the Reading and Literature portion of the OAKS. Also, I addressed how specific demographic indicators contributed to predicting future performance on the statewide assessment.

CHAPTER II

LITERATURE REVIEW

The prior chapter discussed the appropriate use of and necessity for large-scale assessments. While schools and districts are held accountable based on the results of these assessments, statewide assessments are inadequate for informing instruction. The purpose of large-scale measures is to address outcomes, meaning they assess student learning after it has occurred. Because of this limitation of large-scale assessments, alternative measurement systems must be available that are sensitive enough to measure learning on an ongoing basis and predict future performance on outcome assessments. CBMs can often fill the role of assessing student learning in order to inform instruction.

Schools are charged with developing student skills and knowledge. As described in the previous chapter, the stakes have increased to the extent that schools, districts, and states are held accountable for student learning in unprecedented ways (NCLB, 2002). Schools and districts are held accountable for student achievement regardless of real or perceived barriers, including labels of *at-risk* status in the areas of (a) disabilities, (b) limited English proficiency, (c) racial or ethnic minority, or (d) socioeconomic status. Because of the high stakes, teachers must have accessible tools for measuring student performance at regular intervals that inform instruction. The best method for collecting information about student performance in a timely manner is through CBM (Deno, 2003). With the use of CBM, teachers have the opportunity to

track student progress on essential skills and adjust instruction when students do not acquire the skills in a specified amount of time. CBMs may be an incredibly valuable tool because teachers, schools, and districts cannot afford to wait for the results of large-scale, outcome assessments to inform them about student learning. Waiting would not allow teachers to adjust instruction to increase student learning; outcome assessments can only measure what has been learned. When CBM is used as part of an instructional program, especially in early reading, student achievement often improves (Kaminski & Good, 1998).

CBM to Inform Instruction and as a Predictor

In addition to using CBM to progress monitor student learning and to inform instruction, CBM can be used to predict success on large-scale assessments. This can be an important aspect of the usefulness of CBM because it allows student skills to be measured in comparison to outside (more objective) standards. If CBMs can be used to guide instruction and predict, with a high degree of accuracy, whether a student will meet the criteria connected to the large-scale outcome assessment, then it can function as a valuable instructional tool. In the following sections I will discuss how CBM can contribute to instruction through (a) historical perspectives, (b) current accepted uses, (c) the use of oral reading fluency, and (d) CBMs measuring a progression of skills.

CBM to Inform Instruction

The valid interpretations of large-scale assessment results include informing stakeholders about student performance, program evaluation, and program accountability. Based on current research and practicality, informing instruction at the

classroom level is not a valid use of large-scale assessments (Tindal, 2002; Kane, 2002). Assessments useful for informing instruction are considered formative assessments, such as CBM (Deno, 1985; Fuchs & Fuchs, 2004). CBM is a broad term used for a variety of formative assessments used to measure the progression of skills in a particular subject area (Deno, 2003). CBM as an informative counterpart to standardized assessments has been most widely studied in the area of reading, and more specifically in the area of reading fluency (Fuchs et al., 2001). Fuchs et al. (2001), Deno (2003), and others have informed the reading and research community about the uses of CBM.

To better understand why and how CBMs are used to inform instruction requires a look at (a) historical perspectives and critical features of CBM, (b) current accepted uses of CBM, (c) oral passage reading fluency as a measure of overall reading skills, and (d) how CBMs measure a progression of skills. These topics give insight about the origins of CBM and how the measures matured over time. By exploring the history and some of the technical aspects of CBM, a better understanding of how the measures inform instruction and other expanded uses will emerge.

Historical perspective and critical features of CBM. Prior to the widespread use of CBM following groundbreaking work by Deno, Mirkin, and Chiang (1982); Fuchs, Deno, and Mirkin (1984); Marston (1989); and others, the dominant strategy that teachers used for progress monitoring was a strategy called mastery measurement. This approach utilized criterion-referenced assessments and relied on students reaching mastery of particular skill, then assessing the student on the next skill in a progression

(Fuchs & Fuchs, 2004). This approach was riddled with psychometric issues, which led to the development of CBM. The measures were initially developed as a tool for special education teachers so they could evaluate the effectiveness of instruction, document student progress, and adjust instruction (Deno & Mirkin, 1977). As specific measures of reading, writing, and spelling went through further development, the generalizability for CBM was realized. Since the early 1980s, a strong research-base has continued to develop to support the technical adequacy of CBMs and their uses (Deno, 2003).

As stated previously, CBM research initially focused on testing the effectiveness of interventions in special education settings (Deno et al., 1982). The idea was for teachers to improve instruction based on formative evaluation of specific skills. As research progressed, the concept broadened to the idea that measurement systems can use assessment materials selected directly from the instructional programs, hence the term *curriculum-based*. The more specific concept of CBM refers to standardized measurements with specific characteristics. In order for an assessment to be considered a CBM, Deno (2003) stated that it should (a) be technically adequate, (b) include standard measurement tasks, (c) use prescriptive stimulus materials, (d) include administration and scoring guidelines, (e) provide performance sampling procedures, (f) include multiple equivalent samples, (g) be time efficient, and (h) be easy to teach.

When the critical features exist within an assessment system, the utility increases dramatically. Deno (2003) argued that the attributes of CBM provide the opportunity to utilize measures for a wide variety of uses in a large number of settings with a variety of populations. CBM is most widely used as an assessment of basic

reading skills, but current research demonstrated the use of CBM in other subject areas, such as social studies (Espin, et al., 2005), mathematics (Helwig, Anderson, & Tindal, 2002; Clark & Shinn, 2004), and written language (Espin & Tindal, 1998). In addition, the utility of CBM for populations other than special education students is expanding (Deno, 2003).

Current accepted uses. Over the past two decades, the use of CBMs has been refined. As the technology progressed, the use for CBM has expanded. Reading CBMs are currently used as (a) benchmark assessments to screen student skills (Good et al., 2001), (b) a diagnostic assessment tool (Deno, 2003), (c) a progress monitoring tool (Deno & Mirkin, 1977; Hintze & Silberglitt, 2005), and (d) predictors for success on large-scale, outcome assessments (Baker et al., 2008; Missall et al., 2007). The scope of CBM is far beyond just reading fluency. Current developments in reading CBM have expanded to vocabulary and comprehension (Espin et al., 2005; Yovanoff et al., 2005). Also, promising research has been conducted regarding the use of CBM in content area courses and content specific vocabulary (Espin & Deno, 1994-95; Espin, Busch, Shin, & Kruschwitz, 2001).

Deno (2003) stated further valid use of CBMs included determining eligibility for special education, evaluating pre-referral interventions, evaluating instruction, evaluating student reintegration, measuring growth in content areas, and assessing growth in early childhood programs. Furthermore, Fuchs et al. (1984) found that simply using frequent CBM influenced pedagogical practices and student achievement; the use of general CBMs had a positive effect on student learning. With this evidence and

several other studies on CBM (e.g., Missall et al., 2007; Fuchs et al., 2001; Deno, 2003), it is clear that CBMs have vast utility.

As noted earlier, the most widely researched and accepted area in CBM is in basic reading skills (Deno, 2003). Because reading skills can be measured independently and they are progressive in nature, CBM works well as an early reading assessment (Deno et al., 1982). Measuring skills like phonemic awareness and sound identification are accurate measures of pre-reading skills (Kaminski & Good, 1998). According to Kaminski and Good, early reading skill acquisition can be assessed using sound identification, oral blending skills, and fluency. These measures of skill acquisition continue through the more advanced measures of reading skills including passage reading fluency, which is highly predictive of overall reading proficiency (Fuchs et al., 2001).

Oral passage reading fluency. In order for reading curriculum and instruction to be most effective, there must be a continual interplay between reading assessment and instruction (Deno, 1985; Good et al., 2001). Formative assessments are designed to inform educators about student skill development in the context of the skill expectations. Instruction can then be adjusted based on the results of the assessments when measures are used as benchmark and progress monitoring assessments. The most accepted formative measure of reading proficiency is currently oral passage reading fluency (PRF). This measure is widely used by K-12 education because of the large research-base and its practical use as an efficient tool.

Measuring fluency skills is based in the concept that automaticity of lower order skills allows room for higher level cognitive functioning (Samuels, 1979). Studies have associated oral reading fluency with vocabulary (Joshi, 2005) and comprehension skill acquisition (e.g., Baker et al., 2008; Good, et. al., 2001). Assessing fluency is not just measuring the speed at which a student reads; fluency measures the more complex skills of fluidly decoding words and orally forming sentences with prosody (Adams, 1990). The acquisition of these fluency skills is an indication that students continue to develop reading proficiency.

Using CBM to measure a progression of skills. A large portion of the research community has embraced the idea that reading is a progression of skills (Adams, 1990). The NRP's five components of reading are accepted as the building blocks that make up a proficient reader. This is reflected in commissioned work by the federal government (e.g. Adams, 1990; Commission on Reading, 1985; National Institute of Child Health and Human Development, 2000). Also, wording in recent national legislation further supports the claim that reading as a progression of skills has been widely accepted (NCLB, 2002; Individuals with Disabilities Education Act [IDEA], 2004). Furthermore, curriculum and assessment developers make certain that their products reflect the notion that reading is a progression of skills starting with phonemic awareness and phonics, building up to fluency, and finally vocabulary and comprehension skills.

Phonemic awareness, phonics, and fluency are measured as foundational skills to track student progress as beginning readers using CBMs. The assessments measure discrete skills that relate to and predict successful readers (Good et. al, 2001).

According to Good et. al., mastery of rudimentary skills predicts mastery of more complex skills. These researchers argued that students who met the spring kindergarten benchmark on an assessment that measures phonemic awareness, had a significantly higher chance of reaching the winter first grade measure of phonics (nonsense word fluency). This progression continued as students who met the phonics first grade benchmark had a significantly better chance of meeting the spring first grade oral reading fluency benchmark. Furthermore, third grade students who met the third grade ORF benchmark had a significantly better chance of meeting the benchmark on the statewide reading assessment. This outcome presents evidence that ORF is a relatively accurate measure of comprehension (Burns et al., 2002).

Reading CBM as a Predictor of Large-Scale Outcome Measures

In addition to informing teachers about student progress on specific skills, the results of the CBMs must also provide information about the likely outcomes of student performance on large-scale assessments. In other words, student performance on formative assessments, specifically reading fluency CBMs, should have high predictive validity for performance on large-scale statewide reading assessments. If this connection is not present, then the value of reading fluency as a CBM would be reduced; the CBM would not be informing the proper instruction according to the statewide assessment, which is the measure by which schools and districts are deemed effective or not (NCLB, 2002). This logic heavily depends on the technical soundness of the large-scale assessment in question, in that the statewide assessment measures what it purports to measure and that it measures skills associated with necessary skills

to be successful (i.e., internal and external validity). In the cases where the technical adequacy of the large-scale assessment comes into question, the point may be moot. Unfortunately, whether the validity of the large-scale assessment is clear or not, educators are held responsible for student performance on these assessments (NCLB, 2002).

Good et al. (2001) presented evidence that early reading fluency measures related to reading skill progression. The study demonstrated that student proficiency on earlier skills predicts successful mastery of the next skill in the sequence. For example, 90% of the students who met the winter first grade goal for nonsense word fluency (NSF) (measuring pre-reading and early reading skills), met the following spring first grade ORF goal. Also, just 9% of students who did not meet the NSF goal met the spring first grade ORF goal. This sequence of predictions started with pre-reading measures in Kindergarten and continued through third grade ORF scores predicting performance on the Oregon Statewide Assessment Test (OSAT). For example, Good et al. reported that over 96% of third grade students who read 110 correct words per minute (CWPM) met third grade expectations on the OSAT. In contrast, just 9% of the students who read below 70 CWPM met expectations on the OSAT. While the purpose of the Good, et al. study was to establish reasonable cut scores for the early reading measures that are predictive of high-stakes reading outcomes, the results also demonstrated a direct relationship between oral passage reading fluency and performance on large-scale reading assessments that measure comprehension skills.

In similar research, Shaw and Shaw (2002) compared performance on third grade ORF and third grade Colorado State Assessment Program (CSAP). The number of students for this study was small ($n=52$), but the scores for the spring grade 3 ORF had a .80 correlation with performance on the CSAP. They also reported that 90% of the students who read 110 CWPM or more had a 90% chance of meeting proficiency on the CSAP. In two studies with much wider scopes, in terms of number of students (1,766 and 5,472) and grade range, similar results were found (Hintze & Silbergitt, 2005; Silbergitt, et al., 2006). Hintze and Silbergitt used longitudinal data to demonstrate the relationship between reading fluency and performance on a statewide assessment. Hintze and Silbergitt provided correlation coefficients for student performance on two ORF measures (e.g., winter grade 1 ORF and spring grade 1 ORF). The correlation coefficients were documented through spring of grade 3, then a correlation was determined between each of the ORF scores and scores on the Minnesota Comprehensive Assessment (MCA). As previous studies demonstrated, correlations between ORF and the statewide assessment were higher when the assessments were administered in closer temporal proximity. For example, the correlation between the winter grade 1 ORF and the grade 3 MCA was .49, but the correlation between the spring grade 3 ORF and the MCA was .69.

The following studies will demonstrate that over the past several years, research has further substantiated the theory that ORF is a valid predictor of outcome measures and expanded scope of the theory. Wood (2006) investigated classroom and grade level variation in the relationship between ORF and the Colorado Student Assessment

Program (CSAP) and analyzed the predictive validity of ORF and the previous year CSAP score. Wood reported correlation coefficients similar to previous studies (Grade 3, .70; Grade 4, .70; Grade 5, .68), but the study expanded on the correlation coefficients. Statistical analyses of student performance on ORF and CSAP across classrooms and grade-levels demonstrated that the variance in the relationship between the two measures was significant. These analyses suggested that the environment in which students receive instruction might influence the predictive validity of ORF. The analysis of previous year CSAP and ORF showed that ORF scores accounted for additional variance when a multiple regression analysis was conducted using previous year CSAP scores, current year CSAP scores, and current year ORF scores.

Silberglitt, et al. (2006) found a high correlation (.71) between performance on ORF and the MCA at grade 3. The MCA was administered in the spring of the school year and the ORF was given a minimum of five times throughout the school year. Each of the ORF scores was correlated with performance on the MCA. The focus of this study expanded on the study by Wood (2006) by reporting on correlations for grades 5, 7, and 8. The correlation coefficients were much higher in grades 3 and 5 (.71 and .68) than in grades 7 and 8 (.60 and .51). The correlation coefficients decreased in the upper grades, but a moderately strong correlation between the two measures was still evident.

More recently, a study by Schilling, Carlisle, Scott, and Zeng (2007) reinforced the association between ORF and large-scale assessments. This study, with a relatively large number of students (approximately 7,500), reported a range of correlations between ORF in second and third grades (e.g., .65 to .75) and performance on the Iowa

Test of Basic Skills (ITBS) in the same grades. Again, the highest correlations were in the earlier grades when the ORF and large-scale assessments were administered during the same time period (spring ORF and spring ITBS). These correlations were further substantiated with ORF comparisons to the SAT-10 (Baker et al., 2008). Baker et al. found a correlation of .80 between spring grade 2 ORF and a spring administration of the SAT-10. Also, a spring administration of grade 3 ORF and spring administration of OSAT had a .68 correlation. Clearly, the evidence supported earlier assertions of the association between ORF and criterion and norm referenced assessments (Marston, 1989).

The research by Baker et al. (2008) and others demonstrated that passage reading fluency was a strong indicator of more complex skills. Student performance on a short and relatively simple task predicted performance on a time consuming and complex task, namely comprehension skills. Across studies, the results were relatively consistent when tests were administered to elementary aged students around the same time period. The results indicated that the fluency measures were less predictive of comprehension skills when the students were in the upper elementary and middle school grades. This suggested that fluency measures might be less useful as a predictor of comprehension when reading skills are more sophisticated. This provides the impetus for exploring measures that are more predictive of comprehension skills, such as vocabulary CBMs.

Vocabulary Acquisition, Assessment, and CBM

Vocabulary skills are connected to reading comprehension skills (Nagy & Scott, 2000). The difficulty is determining the nature of the connection between the skills. By looking deeper at the specific ways that vocabulary skills and knowledge are acquired and assessed (through different measurement systems, including CBM), the nature of the connection may become clearer. In the following sections, I explore how research addresses vocabulary acquisition, and how vocabulary acquisition is associated with vocabulary assessment practices. I expand on the area of assessment through research associated with vocabulary measurements in specific content areas.

Vocabulary Acquisition and Assessment

Nagy and Anderson (1984) addressed how vocabulary acquisition was a complicated process that develops through a combination of interactions with the written word and formalized instruction. Vocabulary skills were closely associated with reading fluency (Joshi, 2005), reading comprehension (Pearson et al., 2007), and content area knowledge (Espin et al., 2005). Measuring vocabulary skills has also been a complicated process. The format of the assessments and the words used in measurement tools heavily influenced the interpretation and use of assessment results.

Vocabulary acquisition. In recent research, vocabulary acquisition was dependent upon reading skills and frequent interactions with text starting in elementary school (Bear, Invernizzi, Templeton, & Johnston, 2004). Nagy and Anderson (1984) asserted that the best way to develop an adequate vocabulary is to read frequently. The sheer volume of words that proficient readers come in contact with compared to low

skilled readers was staggering. According to Nagy and Anderson, the average middle school student read approximately 1,000,000 words a year (up to 10,000,000 words a year) and came in contact with 3,000 to 4,000 new words during that time. With this sheer number of new words it was necessary for students to read often because “any program of direct vocabulary instruction ought to be conceived in full recognition that it can cover only a small fraction of the words that children need to know” (p. 328). Based on this research and others (e.g. Anglin, 1993; Beck & McKeown, 1999), it was estimated that students learn between 1,000 and 3,000 new vocabulary words per year, most of them acquired outside of specific vocabulary instruction (Nagy & Scott, 2000).

Nagy and Anderson (1984) also reported the number of words that students with proficient reading skills read during a school year was considerably higher than for low performing students. Low performing middle school students may read as few as 100,000 words in a year, while high performing students might read as many as 10,000,000 words in a school year. Based on their logic, this could have drastic effects on vocabulary acquisition. Stanovich (1986) built upon this concept when he described the widening gap between proficient readers and low performing readers as students progressed in the school system. By the time students reach the upper elementary grades, the gap was nearly insurmountable for many students.

Nagy and Scott (2000) introduced aspects of word knowledge that demonstrated its complexity. Incrementality is one of the concepts that had implications for research on instruction and assessment. Pearson et al. (2007) argued that if words were learned incrementally, then it would be useful for assessments to measure the depth of

understanding. This can be accomplished by manipulating the set of distractors and correct responses on multiple-choice vocabulary assessments (Stallman, Pearson, Nagy, Anderson, & Garcia, 1995), but this has not been studied in depth. Another issue that has been addressed is the heterogeneity of vocabulary and word usage. This was based on the idea that words had multiple meanings, depending on context, and the depth of understanding of a word was often dependent on the context (Nagy & Scott, 2000). Further development in vocabulary assessments should be based upon the growing knowledge about vocabulary acquisition and the interrelatedness between vocabulary and comprehension.

Vocabulary assessment. A relatively large amount of research on the progression of early reading fluency skills and how they relate to later, proficient reading skills has been conducted (e.g., Baker et al., 2008; Fuchs, Fuchs, & Compton, 2004; Good et al., 2001; Missall et al., 2007). In contrast, limited research has been conducted related to measuring vocabulary skills (Pearson et al., 2007). These skills are more complex and therefore more difficult to measure. Vocabulary has been measured using matching words to synonyms (Espin et al., 2005), matching pictures to words (Hiebert, 2005), and matching words to definitions (Espin & Deno, 1994-95). The connection between performance on these measures and large-scale assessments is not as widely researched as ORF, but these vocabulary measures have shown promise in expanding CBM to skills that may be more closely associated with more sophisticated reading skills.

The measures used in studies using vocabulary assessments have some shortfalls because of the complex nature of vocabulary. Moreover, it is not completely agreed upon whether vocabulary is a discrete domain or embedded within comprehension (Pearson et al., 2007), but Stahl and Fairbanks (1986), through a meta-analysis, suggested that vocabulary development seems to have a causal role in comprehension. With a growing body of evidence, measuring vocabulary skills may be more efficient than attempting to measure specific comprehension skills using CBM and more useful than ORF as students' reading skills become more complex (Hasbrouck & Tindal, 2006; Yovanoff, et al., 2005; Espin et al., 2005).

Vocabulary skills can be delineated in different ways. Vocabulary words fall into three categories or tiers (Beck, McKeown, & Kucan, 2002). *Tier One Vocabulary* is a group of high-frequency words used commonly in speaking and writing. These words do not necessarily need to be explicitly taught. *Tier Two Vocabulary* includes words that represent an extended, general vocabulary. This tier includes words that are less commonly used to describe common concepts (e.g., astonished rather than surprised) (Pearson et al., 2007). *Tier Three Vocabulary* is a group of specific and specialized words usually reserved for specific content areas (e.g., xylem, mitochondria). The concept of tiers allows for more formulated ways of choosing words for vocabulary measures and allows for a better understanding of the depth of knowledge that a student may have about a word or concept (Pearson et al., 2007). Measuring vocabulary acquisition is very complex and a significant program of

research is likely necessary to gain a better understanding of how to best measure vocabulary skills and interpret the meaning from assessment results (Pearson et al.).

Vocabulary CBM in Content Areas

A series of studies set the context for using vocabulary as a measure of skill and knowledge acquisition at the secondary level (Espin & Deno, 1994-95; Espin & Foegen, 1996). Espin and Deno (1994-95) furnished evidence that measuring specific vocabulary skills could predict success on criterion measures and content knowledge. Espin and Deno (1994-95) extended an earlier study that demonstrated a moderate relationship between oral reading skills and performance on criterion measures for 10th grade students (Espin & Deno, 1993). In the more recent study, Espin and Deno (1994-95) measured vocabulary skills using an assessment that required students to match words to definitions. The correlations between performance on the vocabulary measure and performance on the criterion measures ranged from .40 to .50. Espin and Foegen (1996) produced similar results using vocabulary matching measures and the relationship with three types of criterion measures (comprehension, acquisition, and retention). The vocabulary measures accounted for the largest proportion of variance compared to oral reading and maze measures. The relationship between the vocabulary measure and the criterion measures was moderate to strong ($r=.52$ to $.65$). These studies provided initial evidence that vocabulary measures are valid indicators of knowledge acquisition, but not necessarily gaining overall, more complex reading skills.

Two later studies (Espin et al., 2001; Espin et al., 2005) explored the value of vocabulary CBMs as predicting success using information, knowledge, and skills in the

context of seventh grade content area courses. Espin et al. (2001) used knowledge tests, grades, and performance on the Social Studies portion of the Iowa Test of Basic Skills (ITBS) as criterion variables. The vocabulary measure used as the predictor variable was developed using words and definitions from a seventh grade Social Studies textbook. Alternate forms of the vocabulary measure were given 11 times during the 12-week Social Studies course; the first three were used as the *pre-test* and the last three scores were used as the *post-test*. Students were given a pre- and post-test of the researcher developed content knowledge assessment. The students took the social studies portion of the ITBS after the course ended. The correlation between the vocabulary matching and content knowledge post-tests was over .80. This correlation might be expected because of the nature of the imbedded vocabulary knowledge required for high performance on the knowledge assessment. Furthermore, the correlation coefficients between the vocabulary matching and the ITBS ranged from .56 to .76, demonstrating that there was a moderate to strong relationship between a vocabulary measure and performance on a standardized, large-scale assessment. More interestingly, the correlation between the vocabulary *pre-test* and the knowledge *post-test* was still moderately strong (.66). This demonstrated that student performance on a vocabulary measure predicted student success on an outcome assessment after 12-weeks of instruction.

Espin et al. (2005) used similar vocabulary matching assessments to monitor progress during a 12-week social studies course. Similar to the previous study, the students' skills and knowledge were measured using the ITBS, knowledge assessment,

grades, and a vocabulary measure. The results indicated that there were mean differences and individual differences in growth rates for the vocabulary measure when the assessments were administered on a weekly basis. The differences in growth rates indicated that increased scores on the measure over time demonstrated student learning in course content. Furthermore, the students who made the largest growth rates on the vocabulary measure had the highest grades in the class and the best scores on the knowledge assessment and ITBS. These results reinforced the findings from the Espin, et al. (2001) study. Additionally, Espin's 2005 study expanded the value of vocabulary measures by demonstrating that the measures are sensitive to incremental growth over short periods of time.

Vocabulary as an Indicator of Generalized Reading Skills

In the aforementioned studies, the measurement of vocabulary acquisition was specific to a content area. The students received instruction on specific content and their skills and knowledge about that content area, including vocabulary, were measured using a vocabulary and comprehension assessments. While these studies demonstrated critical aspects of vocabulary CBM, in that they are predictive of future performance and sensitive to growth over short periods of time, they do not necessarily demonstrate that vocabulary CBMs measure more generalized reading skills. According to Pearson et al. (2007), there is a theoretical basis for using vocabulary skills as a measure of overall reading skills. Even though the empirical basis for using vocabulary assessments as a measure of overall reading skill acquisition is limited, Yovanoff et al. (2005) demonstrated that grade-level vocabulary skills (not necessarily associated with a

specific content area) are very closely associated with overall reading comprehension skills.

Theory of Vocabulary as a Measurement of Generalized Reading Skills

White, Graves, and Slater (1990) demonstrated that there are long-term effects that stem from poor vocabulary exposure and acquisition during a child's developmental years. Their research suggested that children with early deficits in vocabulary tend to maintain deficits in overall reading skills over time. Furthermore, the deficits are associated with the differences in the oral and written language to which the children are exposed (Hart & Risley, 1995). The connection between vocabulary and comprehension, especially over the long-term, suggested that further research in the area of vocabulary acquisition and measurement has potentially large ramifications. The connection between comprehension and vocabulary needs more in-depth, technical studies to determine the nature of the connection. Pearson et al. (2007) contended that further research should investigate the effects of vocabulary instruction on comprehension and the transfer of those skills in the short- and long-term. They suggested that the use of computer-based assessments to measure the different domains of vocabulary is a necessary next step in accurate measures of vocabulary acquisition. For example, assessments should be developed that can measure a representative sample of words in a curriculum. Also, assessments that measure estimates of control over specific characteristics may be indexes of vocabulary learning, such as words with common root words (e.g., *equal*) or morphemes (e.g., *ism*); these types of assessments may be a valid measure of overall reading skill acquisition.

Research pertaining to oral, written, receptive, and expressive vocabulary skills suggested that there is a strong connection between vocabulary and comprehension skills. The difficult piece of this connection is the strength of the connection and what aspects of vocabulary are most closely associated with comprehending text. The different assessment methods that have been used and proposed by the research community (see Pearson et al., 2007) attempted to explore the possibilities. The attempts to measure overall reading skills using vocabulary measures have not been fully developed, so the most substantial research has been completed within content area courses (Espin et al., 2005).

The most sizable limitation of the word-to-definition matching formative assessment studies is the context in which the research was conducted. The variables were dependent upon middle and high school content area courses, specifically social studies classes. The amount and nature of vocabulary embedded in social studies courses makes it an ideal context for research on reading and vocabulary skills, but the results of the study are not easily extrapolated to more generalized measures of skill acquisition outside the context of the content area courses. In other words, the vocabulary assessments might have measured knowledge acquisition in the course, rather than acquiring broad reading skills. If true, it does not completely diminish the value of the vocabulary measures as useful formative assessments, but may limit the value of measuring vocabulary skills as a measure of more advanced reading skill acquisition.

Empirical Basis for Vocabulary as a Measure of Generalized Reading Skills

Yovanoff et al. (2005) explored vocabulary assessments as a measure of advanced reading skill acquisition. They addressed the relative value of ORF and a vocabulary measure as a predictor for performance on a comprehension task. The theoretical model they used to explore the value of the two measures was parsimonious, yet applicable to the progression of reading skill acquisition, instruction, and measurement. The model assumed that vocabulary and fluency skills were associated with each other and, at the same time, uniquely contributed to comprehension (to varying degrees depending on grade level and reading ability). They suggested that the relative importance of fluency and vocabulary shifted as students progressed through school. They proposed this model because the focus of reading instruction shifted from *learning to read* in Kindergarten through fourth grade to *reading to learn* in fourth grade and above. With this shift in focus, the complexity of the reading materials increased and formal reading instruction tended to diminish.

Yovanoff et al. (2005) measured oral reading fluency using grade level passages. To assess comprehension skills, students answered 15 literal and inferential comprehension questions after reading a 1,200-word passage. The researchers assessed vocabulary using a synonym matching assessment (in contrast to a word/definition matching assessment or production response). The results indicated that as the grade level increased and the reading skills progressed, the ORF and vocabulary measures were both important indicators of reading skill acquisition, but vocabulary was relatively more important, especially after the fourth grade. In contrast, the correlation

between the vocabulary measure and the comprehension assessment remained steady across the grade levels. This suggested that vocabulary assessments, as a CBM, may be a valid and more stable measure of reading skill acquisition when compared to a reading fluency measure. These findings demonstrated the need for using multiple curriculum-based measures when screening for skill deficits and anticipating future performance on statewide assessments.

Because ORF tends to have diminished predictive validity in the later grades and there is a limited amount of research in the area of vocabulary as a CBM, my research questions are:

1. What is the relationship between the curriculum-based measures of Oral Passage Reading Fluency, Vocabulary measure, and performance on the Reading and Language Arts portion of the Oregon Assessment of Knowledge and Skills (OAKS-Reading) in grade 6?
2. A. What is the relative predictive nature of Oral Passage Reading Fluency and the Vocabulary measure in relation to the statewide assessment for sixth grade students?
2. B. Does adding the non-performance variables of (a) attendance percentage, (b) gender, (c) disability status, (d) English proficiency, (e) ethnicity, and (g) economic disadvantage to the analysis contribute to the predictive nature of Oral Passage Reading Fluency and the Vocabulary measure in relation to the statewide assessment for sixth grade students?

CHAPTER III

METHODOLOGY

This study included data analyses on existing datasets containing results from statewide assessments and district-wide assessments. The analyses only included students who completed all three of the assessments used in the study: (a) Fall PRF, (b) Fall Vocabulary CBM, and (c) OAKS-Reading. Each student had an opportunity to take the statewide assessment three times throughout the school year; the high score was used in the analyses. The specific (a) settings, (b) participants, (c) experimental controls, (d) measures, and (e) data analyses will be described in the following sections.

Setting and Participants

This study was conducted in a school district in the Pacific Northwest with approximately 58,000 residents. In the school district, 10,500 students attended a total of 25 schools in grades Kindergarten through 12th grade. Fourteen elementary schools served students in Kindergarten through fifth grade and two small schools served students in Kindergarten through eighth grade. Approximately 2,400 students attended five middle schools in grades 6 through 8. There are four high schools; two were large, comprehensive high schools, one was a small options program, and one was a small arts focused high school. Approximately 3% of the students in the district accessed services through long-term care and treatment facilities, Education Service District programs, alternative education contracted sites, charter schools, and other out-of-district sites.

Ethnic diversity. The district had relatively low ethnic diversity. Approximately 70% of the students in the district were Non-Hispanic White. Students who were Native American, African American/Black, and Asian made up approximately 2% of the total population, each. Almost 15% of the student population was Hispanic/Latino. There were some differences in the categories across grades and compared to the district totals. For example, Hispanic students made up 16.4% of the sixth grade, while 14.6% of the district as a whole was Hispanic. Statistical analyses were not conducted to determine whether these differences were significant (see Tables 1 and 2).

Gender. Throughout the district and across all grades, girls represented a smaller portion of the students than boys. Girls represented approximately 48% of the student body in sixth grade and across the district. There was a relatively high percentage of girls in the eighth grade compared to the district total and a relatively small percentage of girls in the seventh grade class compared to the district. The sixth grade class had fewer students than both the seventh and eighth grade classes (see Table 1). The following tables supply information regarding the distribution of students across the district.

Table 1 has information pertaining to seventh and eighth grades in order to show that the sixth grade class had similar distributions in the areas of gender and ethnicity compared to the other middle school grades. Table 2 is specific to the population sampled for this study.

Table 1
Percent Gender and Race/Ethnicity for Sixth, Seventh, and Eighth Grades

	Grade		
	6	7	8
Sex			
Girls	48.3	46.2	49.8
Boys	51.7	53.8	50.2
Race/Ethnic Group			
African American	1.2	2.4	1.9
Caucasian	71.1	72.2	74.8
Hispanic	16.6	14.6	12.1
Native American	2.1	1.3	2.8
Asian	1.8	1.3	1.7
Total Number of Students	766	872	834

At-risk. Approximately 16% of the students in the district qualified for special education services. The percent of students in special education, by race/ethnicity was similar to the school district distribution. Native American students and Black students were over represented in special education by approximately 1% and .5%, respectively (see Tables 3 and 4). Student achievement for students with disabilities in the district was low according to statewide assessment results.

Table 2
Percent Gender and Race/Ethnicity in Sixth Grade and District-Wide

	Sixth Grade	District-Wide
Sex		
Girls	48.3	48.1
Boys	51.7	51.9
Race/Ethnic Group		
African American	1.4	1.8
Caucasian	78.1	79.3
Hispanic	16.4	14.6
Native American	2.1	2.4
Asian	2.0	1.9
Total Number of Students	766	10,365

Study participants. The participants in this study included all sixth grade students in the district who were administered the district-wide reading benchmark assessments for oral PRF and Vocabulary CBM in the fall. Also, all of the participants must have taken the OAKS-Reading at least one time during the school year (all students had an opportunity to take the assessment three times). All sixth grade students who participated in all of the three assessments were included in the dataset ($n=678$).

Table 3
Students With Disabilities Within Race/Ethnicity Subgroup

	Primary Race/Ethnicity					Total
	Asian	Black	White	Hispanic	Native American	
Percent in Spec. Ed.	14.9	22.0	16.1	15.6	20.6	16.2
Number of Students	202	186	8,215	1,514	248	10,365

Non-performance indicators also were used in the dataset. Each student who had the necessary scores had three additional pieces of data attached. First, each student had an attendance percentage calculated for the school year. The percentage was determined by the number of days the student attended school divided by the total number of days the student was enrolled in the school district. Second, each student was divided into either male or female based on the information provided in the district's student information system. Finally, students were identified as *at-risk* or not *at-risk*. The *at-risk* category, referred to as NCLB At-Risk, included students identified as special education, English language learner, an underrepresented minority, or economically disadvantaged. The underrepresented minority group includes African American, Latino, and Native American. An a priori decision was made to not include the multi-ethnic, and "declined to answer" in NCLB At-Risk because it was not possible to determine which primary ethnicity with which the students identified themselves. There

were nine students documented as multi-ethnic and three in the “declined to answer” category who were not included in NCLB At-Risk.

Procedures

The scores from the sixth grade passage reading fluency and vocabulary measure were culled from the fall, winter, and spring grade-level benchmarks that were administered to all students in the district in Kindergarten through eighth grade. In every school, the principal and a designated teacher (a special education or Title I teacher in most cases) coordinated the benchmark-testing schedule for each building. All students in the district were tested on all of the grade-level measures, which included individually administered fluency assessments (PRF) and group administered, computer-based assessments (Vocabulary CBM and OAKS-Reading). The fluency assessments included early reading skills and word reading for students in the early elementary grades and passage reading fluency in grades 1 through 8. In most cases, teachers and/or paraprofessionals administered the fluency assessments in a designated area in the school (library, empty classroom, etc.). The vocabulary assessments were administered in a computer lab or in classrooms using a portable laptop computer lab. The schools completed all of the test administration and data-entry within seven student contact days.

Training Procedures

Each assessor was trained to administer the passage reading fluency during two 4-hour sessions conducted by three special education teachers with 10-20 years experience administering standardized and curriculum-based assessments and a school

psychologist with a Ph.D. in school psychology and several years experience in measurement. The assessors were recruited by building principals and were required to have past experience in administering reading fluency measures as part of their job responsibilities. The training sessions provided exposure to the early reading fluency measures, oral passage reading fluency, the on-line CBM assessment system, and ample opportunities to observe and practice administering the assessments. Also, each participant in the training scored each other on the critical elements of administering each assessment using a checklist. The checklist included the following criteria: (a) stopwatch and clipboard ready; (b) read directions verbatim; (c) starts the stopwatch at the appropriate time; (d) mark the last word read at the end of one minute; (e) if the student hesitates for more than three seconds, supply the word and count as an error; (f) put a slash through incorrectly read words; (g) if the student self-corrects, write “SC” and count as correct; and (h) record the total number of words read, subtract errors, and calculates the total words read correctly. The district testing coordinator delivered training sessions for administering computer-based, group administered Vocabulary CBM and OAKS-Reading.

Assessment Administration Procedures

Standard CBM administration procedures were used for the PRF assessments. The assessor was seated at the corner of a table so the student was next to the assessor, but could not readily view what the assessor wrote on the copy of the passage. The assessor had the passage on the clipboard with a stopwatch ready. When the student was seated, the assessor greeted the student and put the student passage on the table in front

of the student. As per the directions on the assessor copy, the assessor pointed to the underlined names in the passage and told the student the names. Then the assessor read the next portion of the directions to the student at the top of the test administrator passage: “I want you to read this story to me. You’ll have one minute to read as much as you can. When I say begin, start reading aloud at the top of the page. Do your best reading. If you have trouble with a word, I’ll tell it to you. Do you have any questions? Begin.”

When the student read the first word in the passage, the assessor started the stopwatch. While the student read the passage, the assessor marked errors by circling omissions and slashing hesitations and mispronunciations. At the end of one minute, the assessor marked a bracket after the last word read and allowed the student to finish the sentence before notifying the student to stop. The assessor documented the total number of words read in a minute and the errors then calculated the correct words read per minute.

The Vocabulary and OAKS-Reading were computer-based, group administered assessments. The vocabulary CBM required the assessor to provide a setting with enough computers for each student (either in a computer lab or a portable laptop computer lab). The assessor had printed directions on the white board in the front of the room. Each student entered a web address, clicked on a large icon labeled “students,” entered their teacher’s name in a text box, selected their own name from a dropdown menu, and selected the *vocabulary* assessment from a dropdown menu. During these procedures, approximately three adults monitored the students to ensure that they

entered proper information. Before the students selected the “take test” button, one of the adults verified the information on the screen. After the students began the assessment, one adult remained in the room during the assessment period until all students finished. The assessment took 10-20 minutes. The assessors assisted the students with navigating the website and any technical problems, but they were instructed to not read any words to the students or provide any word definitions.

Oregon Statewide Assessment

The statewide reading assessment was administered during the spring of the school year. Each student had an opportunity to take the statewide assessment three times, but many students took it only one or two times if they met or exceeded the standard on their first or second opportunity. For the purposes of this study, the students’ highest score was used for analyses.

The students’ teachers proctored the reading OAKS assessments according to the Oregon state assessment guidelines. Each student chose a computer already logged on to the state assessment site. The students selected their names from a dropdown menu and proceeded to take the assessment. Depending on the academic skills of the student and the accommodations necessary (according to an IEP or other individualized plan), the entire test took 45-70 minutes. Each student had the opportunity to take the assessment three times during the school year in order to raise their score from “does not meet” to “meets” or from “meets” to “exceeds.” As noted earlier, the district testing coordinator provided the teachers training regarding proctoring the assessments. Written documentation pertaining to the assessment procedures was provided to the teachers

approximately 2-weeks prior to the first testing opportunity. If the teacher had any further questions regarding the state assessment, the school district assessment coordinator fielded their question via telephone, email, or personal visit.

Measures

To examine how vocabulary and fluency measures were associated with overall reading comprehension, we utilized three different assessments: (a) individually administered, one-minute passage reading fluency (PRF); (b) group administered, computer-based Vocabulary CBM; and (c) OAKS-Reading. I describe the specific administration procedures and technical aspects in the following sections.

Passage Reading Fluency

Using fluency CBMs to measure reading skills has become widely used in the past two decades. While there are a myriad of possibilities for measuring growth over time, many of the assessments have not been developed using sophisticated statistical analyses to determine passage difficulty and equivalency. The passages used in this study were created for the *easyCBM* website and developed using more advanced statistical techniques than simply comparing means and standard deviations when administered to a group of students.

The passages were initially written and revised in an effort to produce 20 alternate forms of grade level passages to be used as progress monitoring and benchmark passages (Alonzo, Park, & Tindal, 2008). The passage developers paid close attention to several criteria while writing the passages. Specifically, each passage (a) tells a story, (b) does not contain dialog, and (c) stands alone with no references to other

passages. Graduate students in the University of Oregon's College of Education wrote the passages. The passages were reviewed for grammar, sentence structure, and grade-level appropriateness by a university professor who is a National Board for Professional Teaching Standards certified English teacher and has a Bachelor's of Arts degree in English. Later, the readability of the passages was determined using the Flesch-Kincaid readability index. Each sixth grade passage had readability between 6.4 and 6.6. Further adjustments were made to the 20 passages so they were similar in format and difficulty. Finally, teachers with a minimum of three years teaching experience reviewed the passages to address grade-level appropriateness. Further analysis of passage equivalency was conducted by administering the passages to groups of middle school aged students. The average correct words per minute and standard deviation informed the researchers about passage difficulty and comparative difficulty. Of the 20 passages, three were reserved for fall, winter, and spring benchmark measures. The 17 remaining passages were retained as progress monitoring measures.

Passage Reading Fluency Administration

Each student was administered the PRF measure three times during the school year (fall, winter, and spring). All students in the district read the same passage within each testing period. The benchmark passages were equivalent according to the analysis described in the previous section. The PRF was an individually administered oral reading fluency measurement. The test administrator read standardized directions to the student informing them of the time limit, that they should do their best reading, and how hesitations will be addressed. The test administrator pointed out and said aloud the

pronunciation of the underlined proper names in the passage. Finally, the test administrator provided an opportunity for questions before the student was told to begin. The student read the passage while the test administrator marked student errors (omissions, mispronunciations, and hesitations) and stopped the student after one minute. Each passage was approximately 300 words in length, so scores could have ranged from 0 to approximately 300 correct words per minute (actual range was 40 to 256 correct words per minute).

In order to establish reliability for the PRF administration, two test administrators scored a portion of the students in two middle schools simultaneously. One was the primary administrator who read the standardized directions. The other was the secondary test administrator sitting behind the primary, listening to the student read, marking errors, and marking a bracket where one-minute ended. A total of 46 students were tested with two trained test administrators present (approximately 20 percent of the assessed sixth graders in two schools). Each test administrator scored the student read passage for rate and accuracy. The scores were compared for reliability on the two domains. The inter-rater reliability was at least 98% across the 46 assessments administered in two separate school settings.

Vocabulary Measure

The vocabulary assessment used in this study treated vocabulary as a discrete skill, measured independently from comprehension. The assessment also attempted to measure vocabulary skills with very little or no context. A large portion of the words in the measurement tool are considered tier two words (Beck, McKeown, & Kucan, 2002).

The words for the measure were pulled from the *World Book Encyclopedia* (2001). From the word list, 60 to 90 items were used in a pilot study. The correct responses for each item were developed from the second word or phrase in a thesaurus for the target word. After the pilot testing, the items that performed adequately were used, which reduced each grade level to a 25-item assessment. The domain that this assessment attempted to measure was *Tier Two Vocabulary* (Beck et al., 2002).

Vocabulary Administration

The vocabulary measure format was a group administered, computer-based word synonym matching assessment. Read aloud was not an option on this measure. A word and three response options were displayed (correct response, near response, and far response) on the computer screen. The choices were a single word or a short, two- to four-word phrase (not word definitions). A short phrase was used when single words in common language were not appropriate synonyms. For example, a test item with the target word of *blunder* had the three possible responses of (a) low hedge (far response), (b) loud noise (near response), and (c) stupid mistake (correct response). The student selected a response and the next item appeared after the student selected the “next” button. The student could select the “back” button at any time during the assessment to change an answer, if desired. In the upper right hand portion of the screen, the student’s progress on the assessment was provided (e.g., #12 of 25). On the last item, instead of a “next” button for the student to click, it said “all done.” When the students were finished, they were provided with a notice thanking them and informing them that they

should notify their teacher that they are finished. Example items from the Vocabulary measure can be viewed in Appendix C.

Reading Portion of the Oregon Assessment of Knowledge and Skills

Each student had the opportunity to take the OAKS-Reading three times during the school year. The assessment was an un-timed, computer-based, multiple-choice test administered to all students in Oregon starting in third grade. The reading assessment had six categories, including (a) vocabulary, (b) read to perform a task, (c) demonstrate general understanding, (d) develop an interpretation, (e) examine content and structure: informative text, and (f) examine content and structure: literary text. On the sixth grade reading OAKS, the distribution of questions in each category was as follows: (a) vocabulary, 20%; (b) read to perform a task, 12%; (c) demonstrate general understanding, 20%; (d) develop an interpretation, 20%; (e) examine content and structure: informative text, 14%; and (f) examine content and structure: literary text, 14% (ODE, 2007). The OAKS assessment system utilized item response theory with a redesigned adaptive algorithm with the goals of improving fidelity to the content, improving fidelity to the test blueprint, and providing as much diagnostic information from the test as possible (ODE, 2007). At the sixth grade level, a score of 222 was considered *meeting the benchmark* and a score of 234 was considered *exceeding the benchmark*.

Reliability analyses conducted by an outside agency for the ODE included the standard errors of measurement for the entire reading assessment and each subtest (strand), which suggested that Oregon's system of assessments provided similar and

consistent information across the range of ability. The standard error of measure (SEM) ranged from 3 to 19 with the largest error at the extreme scores. The SEM at the cut score ranged from four to eight. The OAKS used item response theory in the test administration, which reduced the error in scores. The construct validity studies suggested that the reading OAKS was highly correlated with the California Achievement Test ($r=.75$), the Iowa Test of Basic Skills ($r=.84$), the Northwest Evaluation Association (NWEA) ($r=.79$), and Lexile scores ($r=.76$).

Analyses

The statistical analyses for this study addressed the relative importance of an oral passage reading fluency (PRF) measure and vocabulary measure (Vocabulary CBM) for predicting student success on OAKS-Reading. I provided descriptive statistics including mean, median, and standard deviation of each measure. Also, the analyses determined the correlation coefficients between (a) PRF and OAKS, (b) vocabulary measure and OAKS, and (c) PRF and vocabulary. The descriptive statistics and the correlation coefficient generated from the entire data set (sixth grade student scores) provided a context for the assessments administered to the students and how they related to each other. The correlation coefficients between the PRF and the Vocabulary CBM addressed collinearity issues as well.

Using the same data, I conducted a multiple linear regression analysis to determine which variable accounted for the most variance. The variables included (a) fall PRF, (b) fall Vocabulary CBM, (c) percent attendance, (d) gender, and (e) a combined factor labeled NCLB At-Risk. The NCLB At-Risk factor included students

who (a) have disabilities, (b) have limited English proficiency, (c) were underrepresented minorities, or (d) were economically disadvantaged. These analyses provided information about which of the two CBM scores and nonperformance indicators were most predictive of performance on OAKS-Reading.

CHAPTER IV

RESULTS

Prior to answering the research questions, descriptive statistics for the variables used in the analyses are provided. The first research question was answered utilizing correlation coefficients among the three measurement variables. Also, the two independent, performance variables were analyzed for collinearity issues. The second research question was addressed through two separate multiple regression models. The first analyzed the relative predictive nature of the CBMs in relation to the OAKS-Reading scores and the second model included both measurement variables (CBMs) and three non-performance variables: (a) gender, (b) attendance, and (c) NCLB At-Risk.

Cases Included and General Description

I prepared descriptive statistics for (a) Passage Reading Fluency (PRF), (b) Vocabulary CBM (Vocabulary), and (c) OAKS-Reading (OAKS). Table 5 displays the number of cases, means, standard deviations, minimum scores, and maximum scores. Reminding the reader of the *a priori* decision process, the number of student scores included all students who had scores reported for each of the three measures. A total of 766 students attended the sixth grade in the district during the school year. Of those students, 747 had scores reported for the reading portion of the OAKS. Besides OAKS scores, only students with both CBMs were included. This resulted in a reduction of cases to 684 included in the analyses.

Additionally, an *a priori* decision was made to exclude OAKS scores that fell outside the allowable score range. The Oregon RIT scale ranges from 150 to 300 (ODE Technical Report, 2007). Six students had OAKS scores that fell outside that range. Thus, 678 students were used for this analysis. See Table 5 for complete descriptive statistics.

Table 4
Descriptive Statistics of Assessment Results

Measure	Count	Mean	Std. Dev.	Minimum	Maximum
PRF	678	146.35	37.37	40	256
Vocabulary	678	14.92	4.35	2	25
OAKS-Reading	678	228.32	9.19	204	268

Analyzing for Multicollinearity

Before answering the research questions, it was important to rule out multicollinearity among the variables. Multicollinearity is a situation where there is close to a near perfect linear relationship among some or all of the independent variables in a regression model. In practical terms, this means there is some degree of redundancy or overlap among variables. While multicollinearity is not a fatal flaw, it makes interpretation more difficult. Multicollinearity also causes a loss in power. When there is overlap among some of the variables, it takes more data to disentangle the individual effects of these variables.

I used two tests for determining multicollinearity: (a) correlation and (b) tolerance / Variance inflation factor (VIF). Correlation analysis is the most simple and the least predictive. As a rule of thumb, if the correlation was .90 or larger, the variables would be too closely related to be used in the same regression analysis (Abrams, 2007) and would be presumed to have collinearity.

Table 6 shows that none of the correlations reached the .90 threshold. Correlations ranged between a high of .70 (between OAKS and Vocabulary) to a low of .02 (between Gender and Attendance). Because none of the correlations showed the degree of redundancy or overlap necessary for multicollinearity (Abrams, 2007), all variables were used in the multiple regression analyses. See Table 6 for complete correlations.

Table 5

Zero Order Correlation Matrix

Variable	OAKS	PRF	Vocabulary	Attendance	Gender
PRF	.64**				
Vocabulary	.70**	.56**			
Attendance	.16**	.16**	.10*		
Gender	-.03	-.08*	.06	.02	
NCLB Risk Factors	-.30**	-.21**	-.28**	-.17**	-.03

* p<.05, **p<.01

The second tests were measures of tolerance and variance inflation factor (VIF). “For each independent variable, the tolerance is the proportion of variability of that

variable that is *not* explained by its linear relationships with the other independent variables in the model” (Norusis, 2002, p. 529). According to Tomkins (1992), tolerance values range from 0 to 1. A value close to 1 indicates that an independent variable has little of its variability explained by the other independent variables. A value close to 0 indicates that a variable is almost a linear combination of the other independent variables and would be called multicollinear.

VIF is the second part of this collinearity measure – in fact, VIF is the reciprocal of tolerance in which large values indicate a strong relationship between predictor variables (Mansfield & Helms, 1982). A VIF greater than or equal to 10 suggests multicollinearity (Belsley et al., 1980; Gammie et al., 2003). Again, VIF shows how much of the variance of the coefficient estimate is being inflated by multicollinearity.

The tolerance statistics and the VIF statistics in Table 7 indicate that multicollinearity is *not* a problem for my data in either Model 1 or Model 2 (see Table 7). All tolerances in Model 1 or Model 2 were closer to one, which indicated a lack of multicollinearity (Tomkins, 1992). Tolerance statistics (across both models) ranged from a low of .65 (Model 2 – Vocabulary) to a high of .98 (Model 2 - Gender).

Secondly, all VIF statistics in Model 1 and Model 2 were much lower than 10, which also indicated a lack of multicollinearity (Belsley, Kuh, & Welsch, 1980; Gammie, Jones, & Robertson-Miller, 2003). VIF statistics (across both models) ranged from a low of 1.03 (Model 2 – Gender) to a high of 1.51 (Model 2 - PRF). The complete tolerance and VIF statistics are listed in Table 7.

Table 6

Tolerance / VIF Matrix

	Tolerance	Variance Inflation Factor (VIF)
Model 1		
PRF	.69	1.45
Vocabulary	.69	1.45
Model 2		
PRF	.66	1.51
Vocabulary	.65	1.54
Attendance	.98	1.03
Gender	.95	1.05
NCLB Risk Factors	.90	1.11

Research Question 1: Connection Among Measurement Variables

The first research question analyzed the relationship between student performance on the (a) OAKS-Reading, (b) Vocabulary CBM, and (c) PRF. The relationship was determined by the zero-order correlation coefficients. Table 6 supplies the correlation coefficients for the three measures (all correlations were significant, $p < .01$). The correlation between PRF and OAKS-Reading was strong ($r = .64$) and the correlation between the Vocabulary and OAKS-Reading was strong ($r = .70$). The correlation between the two CBMs (PRF and Vocabulary) was moderate ($r = .56$), but weaker than the correlation between the CBMs and the OAKS.

Question 2A: Predictive Nature of Performance Indicators

The second research question addressed the relative predictive nature of the two CBMs administered. The PRF and the Vocabulary CBM were included in a multiple regression analysis against OAKS-Reading. The ANOVA statistics indicated that one or both of the variables significantly predicted ($p < .0001$) the OAKS-Reading. See Table 8 for the regression summary. Additionally, the coefficients (adjusted $R^2 = .57$) indicated that over 57% of the variance could be explained by PRF and Vocabulary (see table 13).

Table 7

ANOVA Statistics for Model 1, Using PRF and Vocabulary

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	32923.70	2	16461.85	457.74	.000
Residual	24275.49	675	35.96		
Total	57199.19	677			

Table 9 shows results from the multiple regression analysis with OAKS as the constant and PRF and Vocabulary as the predictor variables. Both variables were significant, $p < .0001$. The standardized coefficients indicated that Vocabulary ($\beta = .49$) was relatively more predictive than PRF ($\beta = .37$).

Table 8
Regression of OAKS-Reading on CBMs

Model	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	Sig.	
	<i>B</i>	Std. Error	Beta			
(Constant)	199.65	.99		200.92	.000	
	PRF	.09	.01	.37	12.14	.000
	Vocabulary	1.04	.06	.49	16.21	.000

Table 10 provides further information pertaining to the regression analysis. The semi-partial correlations are included. The semi-partial correlation for Vocabulary (.41) was larger than the semi-partial for PRF (.30). The square of the coefficients showed that 17% of the variance can be uniquely explained by Vocabulary CBM. The PRF measure uniquely explains 9% of the variance. See Table 10 for complete semi-partial correlations.

Table 9
Semi-Partial Correlations: OAKS-Reading on CBMs

Model	Correlations	
	Zero-order	Semi-Partial
PRF	.64	.30
Vocabulary	.70	.41

Question 2B: Predictive Nature With Additional Variables

The third research question addressed whether adding the non-performance indicators to the multiple regression model accounts for more of the variance. The non-performance indicators included gender, attendance, and an NCLB At-Risk. The specifics of each of these factors are provided in the Methods section (see page 63).

Table 6 displays the zero order correlation coefficients for all of the performance and non-performance indicators. There was a negative, weak correlation between Gender and PRF ($r=-.08$). None of the other correlation coefficients related to Gender were statistically significant. Attendance had a positive, weak correlation with all three of the measurement variables. NCLB At-Risk had a negative, weak correlation with all of the performance and non-performance variables, with the strongest correlation coefficients associated with OAKS ($r=-.30$) and Vocabulary ($r=-.28$). See Table 6 for a complete listing of all correlations.

Table 11 supplies information pertaining to the multiple regression model with the additional, non-performance indicators. With the additional independent variables, the two performance indicators were still the most predictive with the Vocabulary CBM slightly more predictive than the PRF measure. Attendance ($p=.09$) and Gender ($p=.17$) were not statistically significant. The NCLB At-Risk factor was significant ($p<.001$), but the standardized coefficient ($\beta=-.08$) indicated that it is far less predictive than the two performance variables. The unstandardized coefficient for the at-risk factor ($B=-1.62$) showed that when the two measurement variables are held constant, the students

in NCLB At-Risk score lower than students not at risk. While the margin is small, it is statistically significant.

Table 10
Regression of OAKS-Reading on CBMs and Other Factors

Model	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	Sig.
	<i>B</i>	Std. Error	Beta		
(Constant)	193.55	5.14		37.70	.000
PRF	.09	.01	.35	11.41	.000
Vocabulary	1.00	.07	.48	15.42	.000
Attendance	9.11	5.35	.04	1.70	.089
Gender	-.64	.46	-.04	-1.38	.170
NCLB At-Risk	-1.62	.51	-.08	-3.20	.001

Table 12 shows the semi-partial correlations associated with all five independent variables. The semi-partials indicate that the PRF (.28) and Vocabulary (.38) accounted for more of the variance, uniquely, than the other variables. Squaring the semi-partial correlation coefficients reveals that Vocabulary accounted for over 14% of the variance and PRF accounted for approximately 8% of the variance. Additionally, NCLB At-Risk accounted for .6% of the variance.

Table 11
Semi-Partial Correlations: OAKS-Reading on CBMs and Other Factors

Model	Correlations	
	Zero-order	Semi-Partial
PRF	.64	.28
Vocabulary	.70	.38
Attendance	.16	.04
Gender	-.03	-.03
NCLB At-Risk	-.30	-.08

The R^2 value and R^2 change values presented in Table 13 document the differences between the two models (without and with the nonperformance indicators). The R^2 of the first model (without the nonperformance indicators) is .574, indicating that slightly over 57% of the variance is predicted by the PRF and Vocabulary measures. The R^2 for the second model is .583, indicating that just over 58% of the variance is predicted with all five variables included. The R^2 change is .01.

Table 12
Variance Accounted for Through Each Model

Model	<i>R</i>	<i>R</i> ²	Adj. <i>R</i> ²	<i>R</i> ² Change	Sig.
1. PRF, Vocabulary	.76	.58	.57	.58	.000
2. PRF, Vocabulary, Gender, Attendance, and NCLB At-Risk	.77	.59	.58	.01	.001

Summary

The correlation coefficients indicated a strong connection between the PRF and OAKS-Reading ($r=.64$) and Vocabulary CBM and OAKS-Reading ($r=.70$). The correlation coefficients between the CBMs and OAKS-Reading were stronger than the correlation between the two CBMs ($r=.58$). The first model used for the multiple regression analysis used the two CBMs as the independent variables. The first model revealed that Vocabulary ($\beta=.49$) was more predictive of student performance on OAKS-Reading than PRF ($\beta=.37$). The second model used in the multiple regression analysis included three nonperformance variables (Gender, Attendance, and NCLB At-Risk). When the three nonperformance variables were added, Vocabulary ($\beta=.48$) and PRF ($\beta=.35$) were the most predictive. The at-risk factor was far less predictive than the two performance variables ($\beta=.08$), but still significant.

CHAPTER V

DISCUSSION

The results from my study indicated that there was a strong connection between the formative measures (CBMs) and the large-scale, outcome assessment (OAKS). In the following sections, I (a) review and summarize the analyses presented in the previous chapter, (b) address limitations to this study, (c) connect the findings to previous research, (d) discuss the practical implications of the findings, and (e) provide suggestions for future research.

Summary of Results From Analyses

This study was conducted to contribute further information about CBMs commonly used as indicators of students acquiring reading skills (Deno, 2003). The usefulness of these CBMs increases if the skills measured by the CBMs are associated with student skills assessed by the statewide assessment (Good et al., 2001). The purpose of this study was to (a) demonstrate the connection between the reading CBMs and the statewide assessment, (b) provide information about the relative predictive nature of the two CBMs used in the study (PRF and Vocabulary), and (c) provide information about the relative predictive nature of the two CBMs and three non-performance indicators (PRF, Vocabulary, Gender, Attendance, and NCLB At-Risk).

The Connection Between Reading CBMs and OAKS

The correlation coefficients indicated a strong connection among the three performance variables (the two CBMs and the OAKS-Reading scores). The correlation between Vocabulary CBM and OAKS-Reading was the strongest ($r=.70$), slightly weaker between Passage Reading Fluency (PRF) and OAKS-Reading ($r=.64$), and the correlation was the weakest between the two CBMs ($r=.56$). These results indicated that there is a connection among the three variables and the connection is stronger between the two CBMs and OAKS than between the two CBMs themselves. These data, not only reinforced the idea that the CBMs provided an indication that students acquired more complex reading skills, but they also provided evidence that the Vocabulary CBM had the strongest connection to performance on OAKS-Reading.

The Relative Predictive Nature of Two CBMs to OAKS

The purpose of analyzing the data, using the two multiple regression models was to understand the relative predictive nature of the performance and nonperformance indicators.

The first regression model. The first model included the performance indicators (PRF and Vocabulary CBM). The standardized coefficient for Vocabulary CBM ($\beta=.49$) was higher than the standardized coefficient for PRF ($\beta=.37$). This indicated that both of the CBMs administered in the fall were predictive of future performance on the OAKS-Reading, with the Vocabulary measure slightly more predictive. The square of the semi-partial correlations indicated that Vocabulary accounted for 17% of the

variance uniquely. This was slightly higher than the 10% of the variance that PRF accounted for uniquely.

The second regression model. The second regression model included the two measurement variables and three nonperformance indicators. Gender and Attendance were not statistically significant, so they did not hold predictive value in this model. The two CBMs and the NCLB At-Risk variable were all statistically significant. The standardized coefficient for Vocabulary CBM ($\beta=.48$) indicated that it was relatively more predictive than PRF ($\beta=.35$) and both were substantially more predictive than NCLB At-Risk ($\beta=-.08$). This model indicated that the two performance variables were more valuable indicators of success on an outcome assessment than the other variables in the model. Although far less predictive, the NCLB At-Risk factor was still significant and should not be dismissed.

Later in this chapter, I address the practical use of the CBMs in instructional settings and ideas for future research utilizing the results pertaining to the CBMs and the at-risk variable. However, before I address the findings and the practical use of CBMs, I will detail some major limitations to this study.

Limitations

The limitations in this study were largely associated with the instruments and the population assessed. The major limitations included (a) mortality, (b) the grade level used, (c) standardization and motivation, and (d) lack of consideration for curricular and instructional variability.

Mortality

The mortality associated with this study lies in the fact that the students included in the study were only the cases with scores for all three measurements. This required that the student was administered the (a) one-to-one PRF measure in the fall (administered during the month of September), (b) the computer based Vocabulary CBM in the fall, and (c) the computer-based statewide reading / language arts assessment – OAKS-Reading. This required that the student attended school in the school district during those periods and cooperated with the assessment conditions. While a mortality rate of 88 students was relatively small, the characteristics of the students who did not participate are unknown. Even with this limitation, the number of cases included in this study was high ($n = 678$) and the distributions of scores for all measurement variables closely mirrored a normal curve (See Figures 1, 2, and 3 in Appendix A). The histograms displayed the frequency of each score on the three measures with a normal curve as a reference. The OAKS (Figure 1) and Vocabulary (Figure 3) distributions more closely mirror the normal curve compared to PRF (Figure 2). In Figure 2 the PRF distribution had a relatively higher kurtosis, with a larger number of students receiving the median score.

Grade Level

This study included students in the school district enrolled only in the sixth grade. The sampling plan was used to decrease potential confounds that could arise when using data from across grade levels. While providing control, using only one grade level reduced the generalizability to other grade levels. The results of this study

pertaining to the predictive nature of the two CBMs and the at-risk factor can only be generalized to other sixth grade students that mirror their demographic variables. More specifically, the results can only be generalized to sixth grade students in similar districts. While this limitation restricts the external validity, it creates opportunities for future research across grade levels and in school districts with a more diverse student enrollment.

Standardization and Motivation

Standardization. Personnel were trained on the administration of the three measures. Every teacher who administered the statewide assessment read the administration manual and listened to a 45-minute presentation on administering OAKS. Generally, teachers proctored the assessments with essentially no other oversight. The assumption was that the licensed staff members abided by the administration rules. Administration of the PRF is a 1:1 setting (as described in the Methods section) and the test administrators were trained how to administer that assessment (also covered in the Methods section). The Vocabulary CBM was computer-based and administered in a group setting, much like the statewide assessment. The licensed teacher proctored the assessment, but there was generally no other supervision. In all of these cases, the assumption was that the administration guidelines were followed. This limitation is present in all studies where measures were administered to large numbers of students. Presumably, large skews in scores would be present if the teachers dismissed the test administration guidelines. However, because those large

skews were not visually present one may reasonably assume that the standardized administration procedures were followed – but that claim cannot be verified.

Motivation. In the same area of concern, student motivation may be a factor (Wise & DeMars, 2005). In sixth grade, neither the reading CBMs nor the OAKS were considered a high stake assessment from the student perspective. With this being the case, student engagement in the activity could be diminished. This was least likely in the PRF measure because of the nature of the administration; it was a one-to-one setting and the assessment only lasted for one-minute. In the case of the Vocabulary measure, motivation may have been a factor because of the difficulty of the words for lower achieving students. Finally, the OAKS may have been the most likely place for diminished motivation. The OAKS required large amounts of decoding text and reasoning. In addition, the OAKS can take some students up to two-hours or more to complete. While this study did not measure student motivation, there were no anecdotal reports of engagement problems present within the district's assessment system.

Measure of Instruction

The final limitation in this study addresses the fact that this was a measurement study based on extant data. The design of the study did not account for differences in instructional approach, curriculum selection, other school-site data, or teacher credentials. The focus of the study was only to look at the connection among student scores on three assessments and three other, nonperformance variables. The nature of the schools and classrooms were not taken into consideration nor measured. In fact, attending school and student schedules were not part of the dataset used for analyses. It

would be naive to think that instructional variability would not impact the connection among the variables.

It is possible that certain aspects of instruction that students were exposed to during the school year confounded the results. In some settings, the instructional approach used by a teacher might have been based on perceived deficits according to the results of the CBMs. If this was the case, where teachers adjusted interventions based on the results of formative assessments, the predictive nature of the CBMs could be diminished. In these classrooms, students who scored low on the fall CBMs might have differential growth in skills over the course of the school year and score better than expected on the OAKS. This must be an area of future research, not only in the middle school grades, but at in the earlier grades as well.

Findings

Schools and districts are held accountable for student achievement through summative assessments and held to a higher standard than in previous years (NCLB, 2002). With the increased stakes of summative assessments for schools and districts, they must have ways to track student learning and adjust instruction based on student performance. CBMs allow for educators to monitor student gains on specific skills (Deno, 2003) and potentially measure student progress toward meeting statewide assessment standards.

PRF as a Predictor

The research over the past two decades provided a strong connection between oral reading fluency achievement assessments in early and intermediate grades (Deno,

2003). Several studies demonstrated the relationship between oral PRF and statewide assessment scores (Baker et al., 2008; Silbergitt et al., 2006; Wood, 2006). These studies reported correlation coefficients between oral passage reading fluency scores and large-scale reading assessments (statewide assessments and SAT-10) in the 3rd grade as high as .80 (Shaw & Shaw, 2002). The other studies reported high correlation coefficients for 3rd grade students ranging from .69 to .71 (Hintze & Silbergitt, 2005; Wood, 2006). My study documented further evidence of this strong connection and extends the strong correlation to sixth grade.

Previous studies suggested a decreased utility in the use of PRF as an indicator of reading achievement as students transition from *learning to read* to *reading to learn* educational environments (Hasbrouck & Tindal, 2006; Silbergitt et al., 2006). For example, Silbergitt reported a correlation of .51 for 8th grade students, a moderate correlation explaining just 26% of the variance. My study reported a correlation of .64, a strong correlation explaining 41% of the variance for sixth grade. While the previous study addressed eighth grade rather than sixth grade students, this is a significant finding; it provides evidence that PRF can be used as a tool by middle school educators that can indicate that a student is acquiring the necessary reading skills to succeed on a high-stakes, summative assessment. My findings supply further evidence that the connection between PRF and summative reading assessments is not drastically diminished by the time the student reaches middle school. Although this study provided evidence that PRF can be used into sixth grade, attempting to measure a more

complicated area, such as vocabulary, may have more predictability in the middle school grades.

Vocabulary as a Predictor

Vocabulary instruction and acquisition are very important for students as they progress through the educational system (Pearson et al, 2007). In addition, vocabulary instruction is complex and vocabulary acquisition is difficult to measure (Nagy & Anderson, 1984; Pearson et al., 2007). Even with the complexity in measuring vocabulary skills, studies have demonstrated the close association between vocabulary acquisition and performance on comprehension assessments (Espin & Deno, 1994-95; Espin et al., 2005). Espin's studies were conducted using content area vocabulary and comprehension measures in the context of a Social Studies courses. The studies reported correlation coefficients as high as .80 between performance on a vocabulary measure and performance on a comprehension assessment (Espin et al., 2001).

The findings in my study extended these findings by providing evidence that the vocabulary assessment and comprehension assessment do not necessarily need to be associated with a content area. The vocabulary assessment used in this study did not focus on any particular subject area and the OAKS-Reading assessed a large range of skills and knowledge. Even with the lack of alignment between two assessments, the correlation was strong ($r=.70$) and the Vocabulary CBM ($\beta=.48$) was more predictive than PRF ($\beta=.37$) providing further evidence that vocabulary acquisition is strongly associated with growth in overall reading skills (Pearson et al., 2007).

Yovanoff et al. (2005) demonstrated that vocabulary was a more stable predictor of performance on a reading comprehension task, across time, compared to oral PRF. Yovanoff et al. used multiple regression models to demonstrate that a vocabulary measure was relatively more predictive of performance on a comprehension measure over time than oral PRF. My study presented further evidence that vocabulary was relatively more predictive than PRF and extends the research by using the Vocabulary CBM to predict performance on the OAKS-Reading rather than a benchmark comprehension assessment that measured a more limited range of skills. Furthermore, the multiple regression analyses used in my study demonstrated that Vocabulary CBM was more predictive than PRF *and* three nonperformance indicators traditionally associated with diminished school success. These results substantiated the claims that several authors have made regarding the strong connection between vocabulary acquisition and overall reading skills (Nagy & Scott, 2000; Espin et al., 2005; Pearson et al., 2007; Yovanoff et al., 2005). The predictive validity of the vocabulary measure in this study provided practical implications for K-12 practitioners and the impetus for further research in the area of more complex reading skills, including vocabulary.

Practical Implications

Schools and districts are held accountable for student performance based on statewide assessment scores, for both the whole student enrollment and for several subgroups (NCLB, 2002; ORS 329.105; ORS 329.488). Because of the legislation associated with increased scrutiny for student achievement, schools and districts must have timely and meaningful data on which educators can base instructional and

programmatic decisions. Previous student scores from statewide assessments are one piece of data that should be used as an indicator of future success on large-scale assessments. However, statewide assessment scores do not supply adequate information pertaining to students' current academic achievement. Statewide assessment scores are not timely (Tindal, 2002) and they tend to assess too broad a range of skills to interpret the data for instructional decision-making (Linn, 2002). Because the statewide assessment is a high stakes assessment for schools and districts and it is inadequate for the purpose of instructional decision-making, the predictive validity of CBMs is that much more important for K-12 educators.

Informed decision-making. Based on the information provided by this and previous studies, school districts should utilize a testing schedule that allows for maximizing the information gained from reading CBMs, especially PRF and Vocabulary. The results of fall CBM administration should allow reading instructors to confidently make decisions regarding instruction and interventions for students who have depressed scores in the sixth grade. Further diagnostic assessments may be needed to accurately determine specific areas of deficit, but low scores on fall CBMs can alert teachers that specific students are at-risk of not meeting benchmarks on the statewide assessment. Also, teachers have an initial indication about students' areas of deficit. If the PRF score is low, the student may need specific instruction to build fluency or decoding skills. Additionally, because the sixth grade passage contains many multi-syllabic words, the student may need specific instruction around decoding the more difficult, information laden words. Using the results of PRF in combination with the

results of the Vocabulary CBM might provide a more in-depth understanding of these deficits. While further analyses of student performance and errors may be required to determine specific intervention types, the information from CBMs is valid and reliable enough to be used as part of a decision-making model for instructional programs at the classroom and building level.

Resource allocation. Building administrators can confidently make decisions about (a) staff allocations, (b) instructional schedules, and (c) curriculum foci based on the results of fall PRF and Vocabulary CBMs in the sixth grade. The findings indicate that students who have relatively low scores most likely need more instruction during the school day that is more directed and specialized. For example, to ameliorate the deficits of the at-risk students principals might commit more staff FTE to specific interventions while altering the school's schedule to accommodate the time necessary for those interventions.

Enrichment. Additionally, students who have higher scores most likely need enrichment and more in-depth instruction. Several years of research provides evidence that vocabulary instruction and practice is necessary for students to develop a substantial vocabulary (Anglin, 1993; Beck & McCowen, 1999; Nagy & Scott, 2000). Research also demonstrated that vocabulary skills are necessary for success in school (Hiebert, 2005; Pearson et al., 2007). The Vocabulary assessment allows school staff to quickly identify students for accelerated programs. Assessments that reveal that specific students may need instruction in particular areas early in the school year, not only can

add efficiency to school and district systems, but can also allow for accelerated learning for all students.

OAKS assessment schedules. Finally, because Oregon allows up to three OAKS assessments per area per year, school officials can use the Vocabulary assessment to determine the most opportune time for testing. Most often, *all* students in a grade-level take the OAKS during the same testing window. Instead, subgroups of students could be provided the opportunity to take the statewide assessment based on their performance on the Vocabulary CBM. For example, students who performed well on the Vocabulary measure early in the school year could take the OAKS during the first opportunity. The students who do not perform well on the measure, could take the statewide reading assessment later in the school year, after more instruction has occurred. Using a test schedule as described could better utilize the Vocabulary CBM as a formative assessment and the statewide assessment as an outcome assessment. This type of test schedule is only realistic if school leaders have a deeper understanding about the connection between the different types of measures *and* the research is substantiated and extended through additional research.

Future Research

The strong correlation between PRF and OAKS reinforces the importance of oral reading fluency. Future research should expand upon the evidence that PRF is closely associated with performance on OAKS even as students enter the middle school. In addition, the high correlation between the Vocabulary measure and OAKS contributes evidence that vocabulary skills are an important aspect of overall reading

skills. These results provide an impetus for future research including (a) studies beyond the sixth grade, (b) the scope of the research pertaining to PRF, (c) the nature of the vocabulary assessment, and (d) separating the NCLB At-Risk factors.

Studies beyond the sixth grade. A strong connection between oral PRF and outcomes assessments in elementary school grades has been established by years of research (Deno, 2003), but the connection tends to diminish in the middle school grades. The results of this study documents that the connection may not be drastically diminished and the connection between the Vocabulary CBM is strong. Similar research should be conducted in later middle school grades. The measures used in this study were developed with close attention devoted to technical adequacy (Alonzo et al., 2008). Because the measures used in this study have documented comparability and grade-level relevance, the connection between PRF and the statewide assessment might be stronger than documented by previous studies (Wood, 2006). Furthermore, the Vocabulary CBM may provide value added predictability later in middle school as it did in for sixth grade students in this study. A large number of students in several districts will allow for the most generalizable information, but determining the strength of the connection between the CBMs and statewide reading assessments in seventh and eighth grade should not require a complex research design. More complicated designs might be necessary to extend PRF research and potentially increase the utility of the measure into the middle school grades.

Future studies involving PRF. Future research should expand upon the findings in this study by looking closer at the complexity of the words and the subject matter in

passages used to measure oral reading skills for students in middle school. Future research could focus on specific attributes of the passages used, possibly using the number of characters per word and number of words per sentence as variables (and documenting student errors). If multiple passages, varying in difficulty, were administered to each student, the additional variables could provide information about how the difficulty of the passages affects the connection between PRF and performance on outcomes assessments. Initial studies in this area could be accomplished by simply administering multiple passages from different grade levels to a large number of students. Then, correlation coefficients could be generated between the PRF for each passage and their performance on the statewide reading assessment. A preliminary study using these methods could supply information about how the difficulty of the passage might influence the predictability.

Future studies involving Vocabulary. Several studies indicated that vocabulary was very closely connected to comprehension within the context of secondary, content-area courses (Espin & Deno, 1994-95; Espin et al., 2005). Also, Yovanoff et al. (2005) demonstrated that vocabulary tends to be a more stable predictor of reading comprehension than PRF, over time. While my study extended this body of research the vocabulary measure in this study used single words in the absence of context. Further research should use a measure that provides the word in context, then requires the student to identify a synonym or definition. While this type of measure would be more complex and would require the student to decode more words in order to answer the question correctly, this type of a measure might give additional information about

students' knowledge of words. Moreover, testing vocabulary in context more closely resembles the real world application that a student would encounter on a day-to-day basis. When words have two meanings, answering patterns could be analyzed to determine what level of understanding students have of the word. This could begin to address the concept of incrementality suggested by Pearson et al. (2007); the items and response options may supply a better understanding of the level of knowledge about specific words that students have acquired.

Future research involving NCLB At-Risk variable. Another area of future research that is necessary pertains to the NCLB At-Risk variable. While this variable was far less predictive than the performance measures, it was still statistically significant. My results yielded evidence that the students in the NCLB At-Risk category scored lower on the OAKS-Reading than students who were not in the *at-risk* category. The NCLB At-Risk category included disabilities, limited English proficiency, race/ethnicity, and socioeconomic status. In future research, each of these categories should be looked at individually. It is possible that one of these categories contributed to scoring lower on the statewide assessment than the others. When other factors are controlled, students classified as having a disability may score lower on the statewide assessment compared to students who are English language learners. Also, subcategories within the major category may need to be a topic of research. For example, students of a particular underrepresented minority group may be more at-risk of low scores than another minority group. Because of the number of variables and categories, this type of research might be very complex.

Research pertaining to the NCLB At-Risk variable might take larger numbers of cases and a more rigorous design. Breaking apart the categories will be difficult because of the variety of disability characteristics, the impact of limited English proficiency, the different race/ethnicity categories, and the gradation of socioeconomic status. For example, disabilities include many types of impairments, including visual, motor, learning, cognitive, sensory, and others. An additional factor could add to the complexity of the research design; students can be classified in more than one category (e.g. have a disability and be an English Language Learner). The research design will most likely need to differentially account for each of the impairments and weight students differently if they fall into more than one category. Furthermore, in order to obtain a high enough number of cases, the research might need to cover large geographical areas, and yet each of these variables and their impact can vary throughout and across regions. While this research might prove to be challenging, it is necessary.

Conclusion

Schools and districts are held to a high standard for student achievement according to statewide assessments scores. States have the authority and obligation to provide sanctions to schools and districts when a high percentage of students do not demonstrate skills on outcome assessments (NCLB, 2002). The expectation is that all students, regardless of barriers based on (a) disabilities, (b) limited English Proficiency, (c) ethnicity, or (d) socioeconomic status, will reach a benchmark standard on the statewide assessment. Because of these high stakes, the validity of CBMs that can inform instructors about student progress toward the standards is drastically important.

Teachers must have efficient and accurate CBMs so they can adjust instructional practices if students are not on track to reach the desired benchmarks.

This study furnished evidence that the vocabulary and reading fluency CBMs were closely associated with and highly predictive of success on the statewide assessment. The connection between PRF and performance on the OAKS was strong, substantiating the use of oral PRF as a measure of reading success in the sixth grade. More importantly, the results demonstrated that the Vocabulary CBM was relatively more predictive of performance on the OAKS than PRF. This gave a strong indication that vocabulary skills were very closely associated with and predictive of overall reading skills.

Practitioners in K-12 education should find the results in this study useful for two reasons. One, the results provided information that can be useful in establishing an assessment schedule that can inform instruction. All students in the sixth grade should be administered CBMs in the area of fluency and vocabulary early in the school year. The results indicated that students who scored low on the CBMs were at substantial risk of not meeting the benchmark on the OAKS-Reading. Therefore, if students were assessed in the fall, teachers had a strong indicator of how the students would perform on the statewide assessment if interventions were not put into place. Second, instructional foci should include fluency, decoding, and vocabulary. Instruction in the sixth grade must include fluency and decoding skills, especially for the students with deficits. Finally, vocabulary instruction and strategies must be a focus in order to increase the chances of students successfully reaching benchmark standards in reading.

APPENDIX A
DISTRIBUTION OF SCORES FOR SIXTH GRADE OAKS, PRF, AND
VOCABULARY

Figure 1

Distribution of Reading OAKS Scores with Normal Curve Reference. The mean, standard deviation, and number of cases are included below the graph.

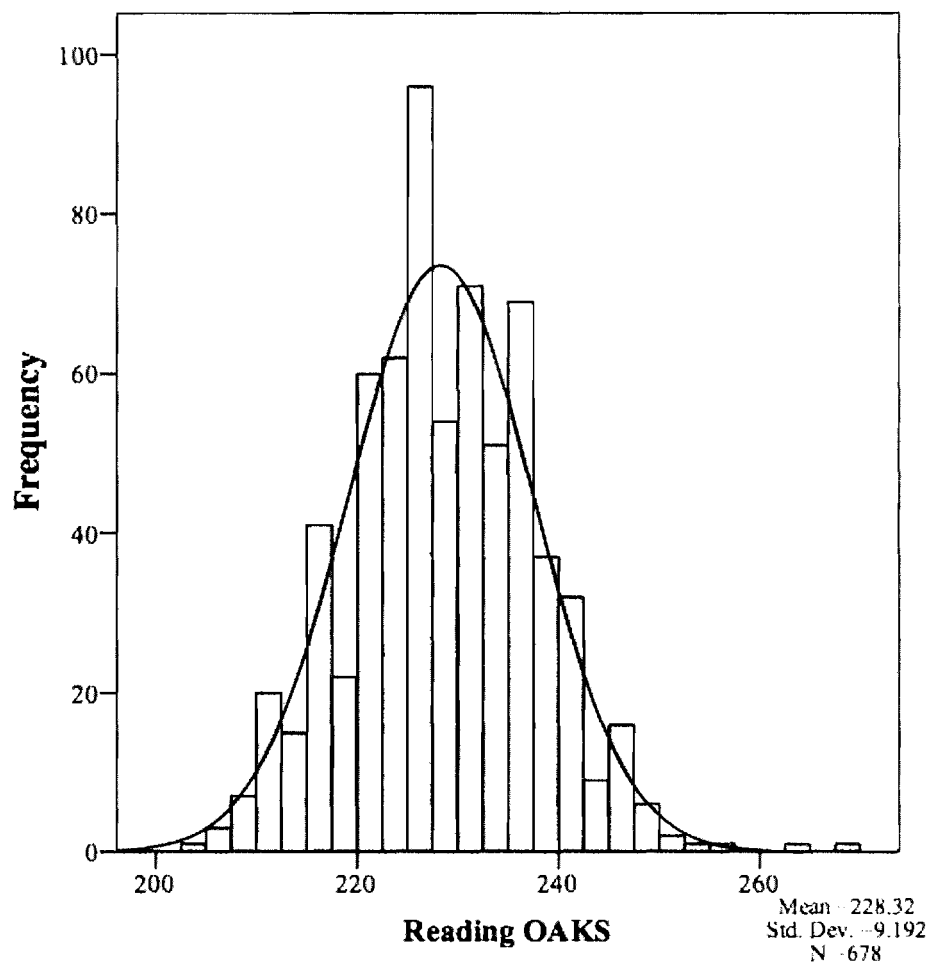


Figure 2

Distribution of PRF Scores with Normal Curve Reference. The mean, standard deviation, and number of cases are included below the graph.

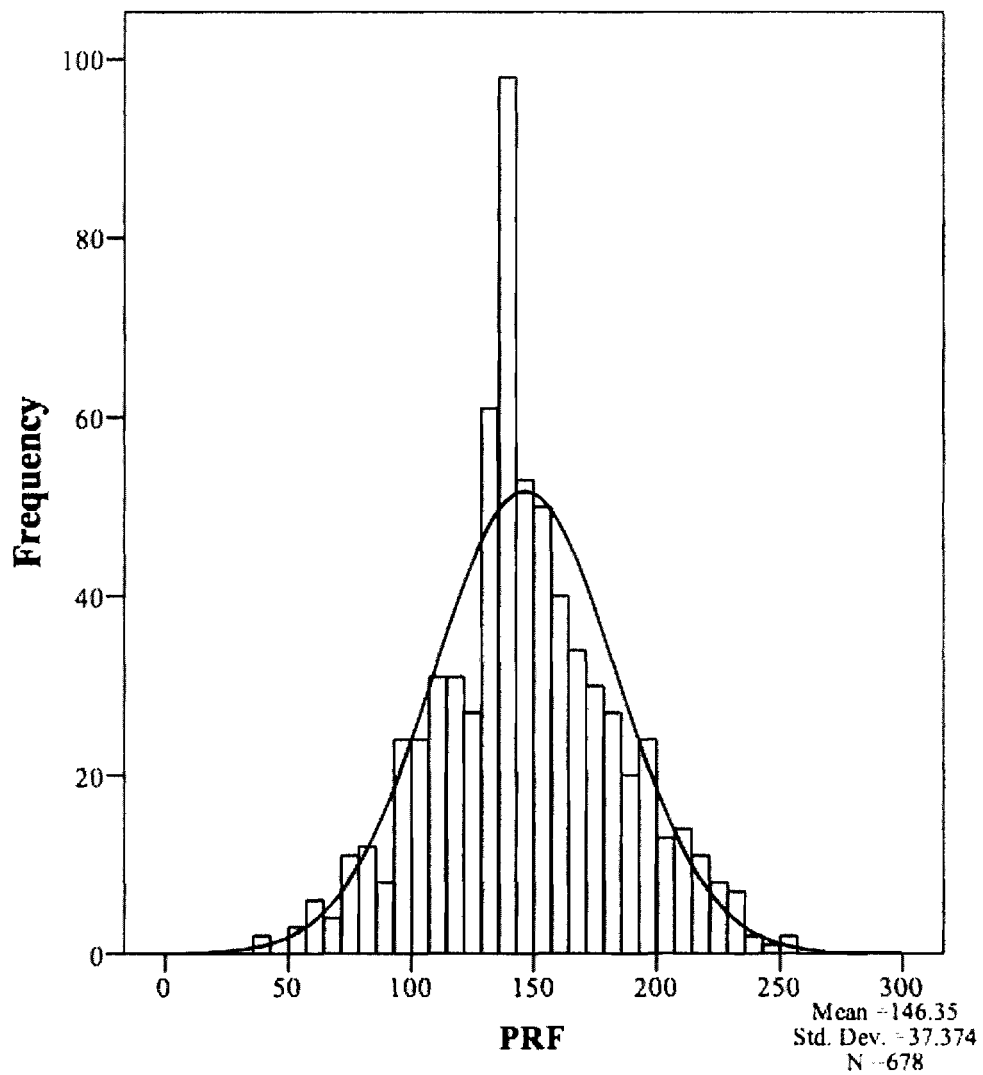
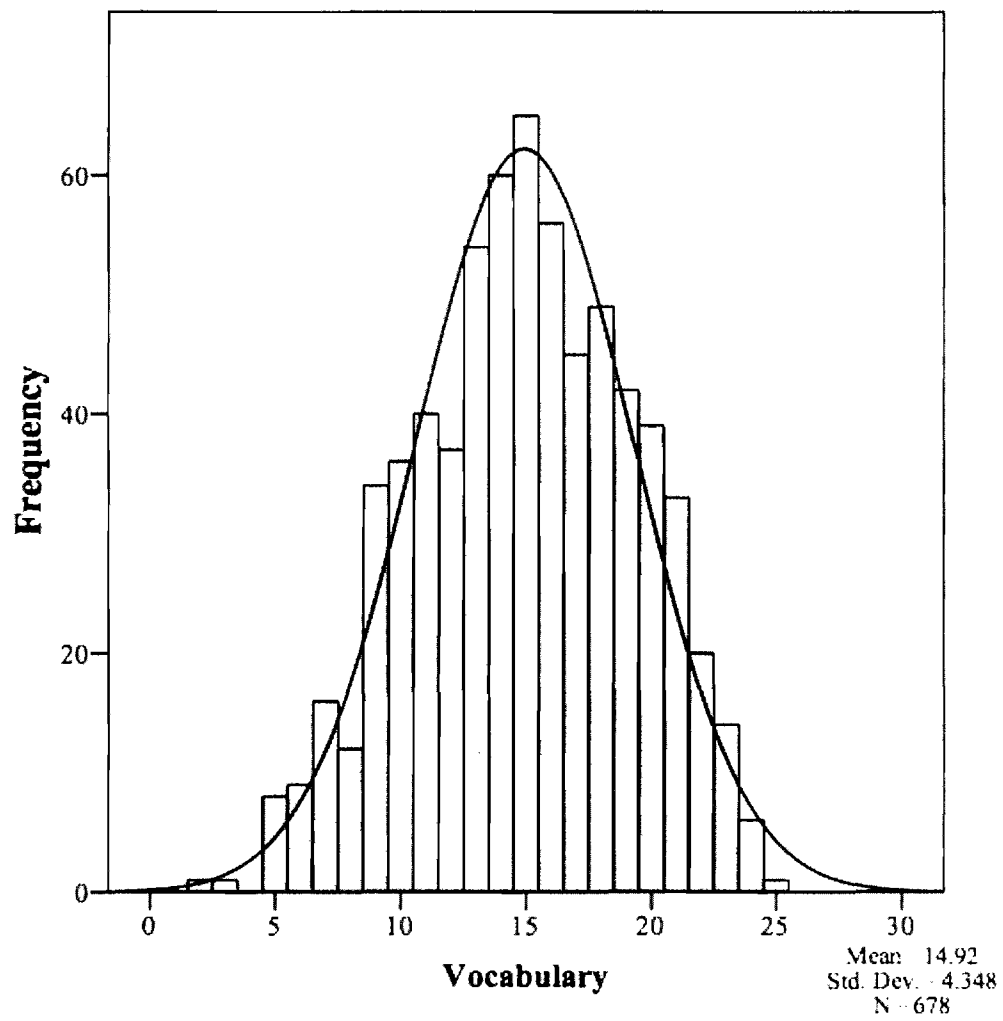


Figure 3

Distribution of Vocabulary Scores with Normal Curve Reference. The mean, standard deviation, and number of cases are included below the graph.



APPENDIX B

FALL SIXTH GRADE BENCHMARK PASSAGE, STUDENT AND
ASSESSOR COPIES

Student Copy Grade 6-Fall

One day, Mr. Johnson assigned his class a group project. He told the students that they could choose their groups and work in groups of three or four. There was a rule that everyone had to do an equal share of the work. After they finished the project, each group member would fill out a form telling Mr. Johnson whether the other group members had contributed equally to the project.

Nancy had mixed feelings about group projects. When group projects were assigned, everyone wanted Nancy to be in their group because she usually got good grades. Nancy liked this feeling of popularity. But, because she was a hard worker, she often ended up doing all of the work on group projects. She was worried that it would be hard for her to fill out the form about contributions at the end of the project.

Nancy decided to talk to Mr. Johnson about her concerns. Mr. Johnson would not give Nancy an option to work by herself. He explained to Nancy that there was a life lesson to be learned. He told Nancy that if she chose the right partners, she wouldn't have to do all of the work and she would have no problem filling out the form at the end of the project.

Nancy thought about what Mr. Johnson said as she looked at the invitations left in her locker. She decided to ignore most of the invitations and choose partners who she knew would be willing to do some work. As soon as Nancy had decided on her group, she met with them and made a plan for how everyone would contribute to the project. Nancy was sure that everyone would do their part and they would get a good grade. She was proud of herself for choosing such a hardworking group.

Assessor Copy Grade 6-Fall

Student Name: _____ Date: _____

1. Place the first passage without numbers in front of the student. Point to any names in the document and tell the student how to say the name. Then say: **"This is a story about Mr. Johnson and Nancy. I want you to read this story to me. You'll have 1 minute to read as much as you can. When I say "begin," start reading aloud at the top of the page. Do your best reading. If you have trouble with a word, I'll tell it to you. Do you have any questions? Begin."**
2. Start the timer.
3. While the student is reading, mark errors with a slash (/).
4. At 1 minute, mark the last word read with a bracket (]).
5. When the student gets to a logical stopping place, say **"Stop."**

<p style="text-align: center;"><u>One day, Mr. Johnson</u> assigned his class a group project. He told the students that they could choose their groups and work in groups of three or four.</p> <p>There was a rule that everyone had to do an equal share of the work. After they finished the project, each group member would fill out a form telling Mr. Johnson whether the other group members had contributed equally to the project.</p> <p style="padding-left: 40px;"><u>Nancy</u> had mixed feelings about group projects. When group projects were assigned, everyone wanted Nancy to be in their group because she usually got good grades. <u>Nancy</u> liked this feeling of popularity. But, because she was a hard worker, she often ended up doing all of the work on group projects. She was worried that it would be hard for her to fill out the form about contributions at the end of the project.</p> <p style="padding-left: 40px;">Nancy decided to talk to Mr. Johnson about her concerns. Mr. Johnson would not give Nancy an option to work by herself. He explained to Nancy that there was a life lesson to be learned. He told Nancy that if she chose the right partners, she wouldn't have to do all of the work and she would have no problem filling out the form at the end of the project.</p> <p style="padding-left: 40px;">Nancy thought about what Mr. Johnson said as she looked at the invitations left in her locker. She decided to ignore most of the invitations and choose partners who she knew would be willing to do some work. As soon as Nancy had decided on her group, she met with them and made a plan for how everyone would contribute to the project. Nancy was sure that everyone would do their part and they would get a good grade. She was proud of herself for choosing such a hard working group.</p>	<p>13</p> <p>28</p> <p>45</p> <p>59</p> <p>70</p> <p>81</p> <p>95</p> <p>109</p> <p>125</p> <p>143</p> <p>144</p> <p>157</p> <p>174</p> <p>190</p> <p>207</p> <p>214</p> <p>227</p> <p>241</p> <p>257</p> <p>273</p> <p>287</p> <p>303</p> <p>305</p>
--	---

Total Words Read: _____ -# of Errors: _____ = CWPM _____

APPENDIX C

FALL SIXTH GRADE BENCHMARK EXAMPLE VOCABULARY ITEMS

drift #1 of 25 wagon harness patchwork quilt wind-piled snow[Next](#)**abolish** #2 of 25 take prisoner do away with eat rapidly[Back](#)[Next](#)**secrecy** #5 of 25 finding out sharing knowledge concealing from[Back](#)[Next](#)**advantageous** #10 of 25 helpful confused errorless[Back](#)[Next](#)

REFERENCES

- Abrams, D. R. (2007). Introduction to Regression Princeton University. Retrieved March 1, 2010, from http://dss.princeton.edu/online_help/analysis/regression_intro.htm
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT.
- Alonzo, J., Park, B., & Tindal, G. (2008). *The Development of Middle School Passage Reading Fluency Measures for use in a Progress Monitoring Assessment System*. (Technical Report #46). Eugene, OR: Behavioral Research and Teaching.
- Anglin, J. M. (1993). Knowing versus learning words. *Monographs of the Society for Research in Child Development*, 58, 176-186.
- Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Thomas Beck, C. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, 37, 18-37.
- Bear D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2004). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Columbus: Prentice Hall.
- Beck, I. L., & McKeown, M. G. (1999). Comprehension: The sine qua non of reading. *Teaching and Change*, 6, 197-211.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: The Guilford Press.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Data and Sources of Collinearity*. New York: John Wiley and Sons.
- Burns, M. K., Tucker, J. A., Hauser, A., Thelen, R. L., Holmes, K. J., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: A potential criterion for curriculum-based assessments. *Assessment for Effective Interventions*, 28, 1-7.

- Clark, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234-248.
- Commission on Reading (1985). *Becoming a nation of readers: the report of the commission on reading*. Washington: The National Institute of Education.
- Conley, D. T. (2007). *Policy analysis: Oregon student assessment systems*. Eugene, Oregon: Education Policy Improvement Center (EPIC).
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184-192.
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based Program Modification: A Manual*. Reston, VA: Council for Exceptional Children.
- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 19* (8), 1-16.
- Elmore, R. F. (2004). The problem of stakes in a performance-based accountability system. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 277-292). New York: Teachers College Press.
- Espin, C. A., & Deno, S. L. (1993). Performance in Reading from Content Area Text as an Indicator of Achievement. *Remedial and Special Education, 14*(6), 47-59.
- Espin, C. A., & Deno, S. L. (1994-95). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20*, 121-142.
- Espin, C. A., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in the content areas: Vocabulary matching as an indicator of progress in social studies learning. *Journal of Learning Disabilities, 38*, 353-363.
- Espin, C. A., Busch, T. W., Shin, J., & Kruschwitz, R. (2001). Curriculum-based measurement in the content areas: Validity of vocabulary-matching as an indicator of performance in social studies. *Learning Disabilities: Research and Practice, 16*(3), 142-151.

- Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content area tasks. *Exceptional Children, 62*(6), 497-514.
- Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Eds.), *Advanced applications of curriculum-based measurement* (pp. 214-253). New York: Guilford.
- Fowler, F. C. (2009). *Policy Studies for Educational Leaders: An Introduction* (3rd ed.). Boston: Pearson.
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.
- Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659-671.
- Fuchs, L. S., & Fuchs, D. (2004). Determining adequate yearly progress from Kindergarten through grade 6 with curriculum-based measurement. *Assessment for Effective Intervention, 29*(4), 25-37.
- 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.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5*, 239-256.
- Gammie, E., Jones, P., & Robertson-Miller, C. (2003). Accountancy undergraduate performance: A statistical model. *Accounting Education, 12* (1), 63-78.
- Good, R. H., Simmons, D. C., & Kame'enui E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.
- Hasbrouck, J., & Tindal, G. (2006). Oral Reading Fluency Norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*, 636-644.

- Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education, 36*, 102-112.
- Herman, J. L. (2007). Accountability and assessment: Is public interest in K-12 education being served? (CRESST Report 728, October, 2007). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hiebert, E. H. (2005). In pursuit of an effective, efficient vocabulary curriculum for elementary students. In Hiebert, E.H., & Kamil, M.L. (Eds.), *Teaching and Learning Vocabulary: Bringing Research to Practice* (pp 243-263). Mahwah, NJ: Erlbaum.
- 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.
- Individuals With Disabilities Education Improvement Act of 2004. (2004). *Public Law #108-466, 614 Stat, 2706*.
- Joshi, R. M. (2005). Vocabulary: A critical component of comprehension. *Reading & Writing Quarterly, 21*, 209-219.
- Kame'enui, E. J., Fuchs, L., Fracis, D. J., Good R. H., O'Connor, R. E., Simmons, D. C., Tindal, G., & Torgesen, J. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher, 35*(4), 3-11.
- Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Eds.), *Advanced applications of curriculum-based measurement* (pp. 113-142). New York: Guildford.
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices, 21*(1), 31-41.
- Linn, R. L. (2002). Validation of the uses and interpretations of results of state assessment and accountability systems. In Tindal, G. & Haladyna, T. M. (Eds.), *Large-scale assessment programs for all students* (p.p. 49-66). Mahwah, NJ: Erlbaum.
- Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *American Statistician 36*(3), 158-160.

- Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Eds.), *Curriculum-Based Measurement: Assessing Special Children*. (pp. 18-78). New York: Guilford
- Marston, D., Fuchs, L.S., & Deno, S. L. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11*, 77-90.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.
- Missall, K., Reschly, A., Betts, J., McConnell, S., Heistad, D., Pickart, M., Sheran, C., & Marston, D. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36*, 433-452.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304-330.
- Nagy W. E. & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. 2, pp. 269-282. Mahwah, NJ: Earlbaum.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: National Commission on Excellence in Education.
- National Institute of Child Health and Human Development. (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* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- No Child Left Behind Act of 2001. (2002). *Public Law #107-110, 115 Stat. 1425*.
- Norusis, M. J. (2002). *SPSS 11.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Oregon Department of Education. (2007). *Oregon statewide assessment system technical report: Volume 1, Annual Report*. Salem, Oregon: Oregon Department of Education, Office of Assessment and Information Services.
- Oregon Department of Education (2009). *AYP Policy and Technical Manual*. Salem, Oregon: Oregon Department of Education, Office of Assessment and Accountability.

- Oregon School Boards Association. (2005). *Education Improvement in Oregon*. Retrieved May 15, 2008, from <http://www.osba.org/covered/curricul/edimprov.htm>.
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*, 282-296.
- Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32*, 403-408.
- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*, 429-448.
- 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.
- Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*, 527-535.
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110.
- Stallman, A. C., Pearson, P. D., Nagy, W. E., Anderson, R. C., & Garcia, G. E. (1995). *Alternative approaches to vocabulary assessment* (Technical Report No. 607). Urbana-Champaign, IL: Center for the Study of Reading, University of Illinois.
- Stanovich, K. E. (1986). Mathew effects in reading: Some consequences of individual differences in the acquisition of learning to read. *Reading Research Quarterly, 21*, 360-407.
- Stallman, A. C., Pearson, P. D., Nagy, W. E., Anderson, R. C., & Garcia, G. E. (1995). *Alternative approaches to vocabulary assessment* (Technical Report No. 607). Urbana-Champaign, IL: Center for the Study of Reading, University of Illinois.
- Tindal, G. (2002). Large-scale assessments for all students: Issues and Options. In Tindal, G. & Haladyna, T. M. (Eds.), *Large-Scale Assessment Programs For All Students* (p.p. 1-24). Mahwah, NJ: Erlbaum.
- Tomkins, C. A. (1992). Using and interpreting linear regression and correlation analysis: Some cautions and considerations. *Clinical Aphasiology, 21*, 35-46.

- White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology, 82*, 281-290.
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*, 1-17.
- Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*, 85-104.
- 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* (3), 4-12.