# PREDICTING OUTCOMES ON HIGH STAKES ASSESSMENTS FOR MIDDLE SCHOOL STUDENTS: A COMPARISON OF CURRICULUM-BASED MEASURES AND EXTANT DATASETS

By

Nathan A. Stevenson

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Special Education – Doctor of Philosophy

#### ABSTRACT

# PREDICTING FAILURE ON HIGH STAKES ASSESSMENTS FOR MIDDLE SCHOOL STUDENTS: A COMPARISON OF CURRICULUM-BASED MEASURES AND EXTANT DATASETS

By

## Nathan A. Stevenson

As a school-wide framework, Multi-Tired Systems of Support (MTSS) relies on the prevention and early identification of students at risk of academic failure (Sugai & Horner, 2009). Approaches to early identification of students in need of support include the administration of universal screening assessments and the analysis of existing student data such as attendance, grades, office discipline referrals, and prior performance on statewide assessments. However, there is little research that directly compares the accuracy and reliability of these approaches, particularly in middle grades. This investigation provides a direct comparison of curriculum-based measures in reading and the examination of archival data at the middle school level for the identification of students at risk for academic failure. Data were collected for students in grades seven (n = 197) and eight (n = 237). Data were analyzed through hierarchical logistic and linear regressions using outcomes on reading subtests of Michigan Education Achievement Program (MEAP) and ACT Explore<sup>®</sup> as the dependent variables. Results inform how data from universal screening assessments and existing sources can be used to accurately and efficiently identify students in need of academic support.

Copyright by NATHAN A. STEVENSON 2015 Dedicated to Jenny, Carter, and Lily. Thank you for everything you are and all that you do. I could not have done this without out you.

# ACKNOWLEDGEMENTS

I wish to thank the incredible group of teachers, administrators, and consultants I work with every day. Special thanks to Dr. Cythia Okolo, for her unwavering support, advisement, critique, and flexibility throughout this experience. Thank you to Dr. Sara Witmer and Dr. Mark Reckase for their superb advice and service on this committee. To Dr. Troy Mariage, who helps me see the bigger picture and keeps me excited about everything. To my teammates Dave Pike and Nate Jarvie for their tremendous contributions to the education of all students. And finally to my wife, Jennifer, whose support and encouragement made this possible.

LIST OF TABLES	vi
INTRODUCTION	1
CHAPTER 1: Rationale	3
Universal Screening	3
Screening in Secondary Schools.	4
Screening in Middle Schools	7
CHAPTER 2: Literature Review	9
Universal Screening Measures in Reading	9
Oral Reading Fluency	9
Maze Reading Comprehension	10
easyCBM Multiple Choice Reading Comprehension	12
Farly Warning Signs	13
Attendance	15
Behavior	15
Course Failure	17
Statewide Deading Assessments	1 /
Pagant Pasaarah in Middla Sahaal Saraaning	20
Posoarch Questions	20
Research Questions	
CHAPTER 3. Methods	24
Participants and Setting	24
Data Collection	24
Reading Screening Measures	26
easyCBM Multiple Choice Reading Comprehension	27
Reading-Curriculum Based Measure	27
Maze Reading Comprehension	$\frac{2}{28}$
Farly Warning Signs	28
Attendance	28
Office Discipline Referrals	$\frac{20}{20}$
Failing Grades	20
Reading Outcome Measures	2)
Michigan Education Assessment Program	29
ACT Explore <sup>®</sup> Description	29
ACT Explore Reading	21
Data Analysis.	32 22
Data Screening.	22 22
Binary Analysis.	32 25
	33
CHAPTER A: Results	36
Data Screening	36
Data Soloming	50

# **TABLE OF CONTENTS**

Demographic Variables	36
Binary Analysis	37
Early Warning Signs	37
Curriculum Based Measures	40
Linear Analysis	42
Early Warning Signs	43
Curriculum Based Measures	44
CHAPTER 5: Discussion	48
Summary	48
Question 1	49
Question 2	51
Question 3.	53
Limitations	55
Implications	58
Suggestions for Future Research	60
APPPENDIX	63
BIBLIOGRAPHY	84

# LIST OF TABLES

Table 1 Summary of Descriptive Statistics and Tests for Normality	64
Table 2 Model Fit for Seventh Grade EWS Variables on MEAP Reading	65
Table 3 Model Fit for Eighth Grade EWS Variables on MEAP Reading	66
Table 4 Logistic Regression Results for Seventh Grade EWS on MEAP Reading	67
Table 5 Logistic Regression Results for Eighth Grade EWS on MEAP Reading	68
Table 6 Model Fit for Eighth Grade EWS Variables on ACT Explore <sup>®</sup> Reading	69
Table 7 Logistic Regression Results for Eighth Grade EWS on ACT Explore <sup>®</sup> Reading	70
Table 8 Model Fit for Seventh Grade CBM Variables on MEAP Reading	71
Table 9 Model Fit for Eighth Grade CBM Variables on MEAP Reading	72
Table 10 Logistic Regression Results for Seventh Grade CBM on MEAP Reading	73
Table 11 Logistic Regression Results for Eighth Grade CBM on MEAP Reading	74
Table 12 Model Fit for Eighth Grade CBM Variables on ACT Explore <sup>®</sup> Reading	75
Table 13 Logistic Regression Results for Eighth Grade CBM on ACT Explore <sup>®</sup> Reading	76
Table 14 Linear Regression Results for Seventh Grade EWS on MEAP Reading	77
Table 15 Linear Regression Results for Eighth Grade EWS on MEAP Reading	78
Table 16 Linear Regression Results for Seventh Grade CBM on MEAP Reading	79
Table 17 Linear Regression Results for Eighth Grade CBM on MEAP Reading	80
Table 18 Linear Regression Results for Eighth Grade EWS on ACT Explore <sup>®</sup> Reading	81
Table 19 Linear Regression Results for Eighth Grade CBM on ACT Explore <sup>®</sup> Reading	82
Table 20 Comparison Data of Enrollment and Academic Achievement for 2012	83

#### **INTRODUCTION**

Multi-Tiered Systems of Support (MTSS) is a framework for the delivery of instruction and intervention to support positive academic and behavioral outcomes. MTSS incorporates both academic and behavior supports (Sugai & Horner, 2009). With the integration of Response-to-Intervention (RtI) and School-Wide Positive Behavior Supports (SW-PBIS), MTSS provides schools with a comprehensive structure of services including: (a) early and accurate identification of students in need of support, (b) delivery of research-based instruction and interventions, (c) systematic collection and analysis of student data, and (d) explicit methods of data-based decision making (Fuchs, 2003; Fuchs & Fuchs, 1998; Fuchs, Fuchs, & Speece, 2002; Fuchs, Mock, Morgan, & Young, 2003; Sugai & Horner, 2009; Vaughn, Linan-Thompson, & Hickman, 2003). While there is much debate on the impact of MTSS on student achievement, it is widely accepted that the early identification and prevention of academic failure is a reasonable, and responsible approach to ensuring that all students meet or exceed the standards for career and college readiness (Fuchs, 2003; Fuchs, Fuchs, & Compton; 2013).

Early identification and treatment of students in need of support is a core function of any MTSS system (Batsche et al., 2005; Hosp, Hosp, & Dole, 2011; Salvia, Ysseldyke, & Bolt, 2009). At the elementary level, identification of learners at-risk for not reaching grade level benchmarks is often done using Curriculum-Based Measures (CBM) such as oral reading fluency and Maze reading comprehension. CBM are a key source of data in (a) screening for risk of academic failure, (b) progress monitoring, (c) program evaluation, (d) diagnosis of specific learning disabilities, and (e) a recommended component of a balanced assessment system (Salvia, Ysseldyke, & Bolt, 2009).

At the secondary level, existing data sets such as state achievement tests and course

grades are used more often as screening tools than are CBM (Heppen & Therriault, 2008). Copious research has demonstrated the predictive power of extant data at the high school level with such factors as attendance, behavior, grades, and results from past state assessments in predicting high school dropout (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Baydar, Brooks-Gunn, & Furstenberg, 1993; Casillas et al., 2012; Heppen & Therriault, 2008; Jerald, 2006; Pinkus, 2008). These data, commonly known as Early Warning Signs (EWS), are used in high schools throughout the United States to serve a screening function similar to that of CBM at the elementary level (Kennellly & Monard, 2007). While screening with CBM reading assessments in elementary is done to predict which students are at risk of not reaching grade level performance benchmarks, EWS at the high school level are typically used to predict which students have academic or behavior problems and are therefore at risk for dropout.

However, despite the growth of research in MTSS, CBM, and EWS, there remains several unanswered questions with regard to the technical adequacy and usefulness of such predictive data at the middle school level. The following chapters discuss the purpose, rationale, and brief history behind the use of CBM and EWS as universal screening measures, a review of existing literature regarding the use of CBM and EWS in middle school, along with methods, results, and discussion of an investigation comparing the use of CBM and EWS in predicting failure on high stakes tests.

# **CHAPTER 1: Rationale**

# **Universal Screening**

To provide effective intervention and supplemental services to students that struggle, schools must first identify those in need of support. Early and accurate identification of students at risk for developing persistent skill deficits provides schools the best opportunity to intervene early and put students back on track to success. One approach to identifying students in need of support is through the use of universal screeners. Universal screeners are brief assessments given to all students in a school as an initial means for determining which students may be in need of additional instructional supports and services (Hosp, Hosp, & Dole, 2011). Universal screeners in schools are analogous to measures of blood pressure or body temperature in a routine office visit to a physician. Such assessments are not diagnostic but are strong indicators that problems exist. Early identification through screening procedures aids in identifying those individuals who, without intervention, would "develop serious and chronic academic problems" (Fuchs, Fuchs, & Compton, 2013 p. 265). Though there is much debate on the efficacy of screening with regard to achieving better student outcomes for those at risk of academic problems, it is widely accepted that a focus on early identification and prevention of academic difficulties is a promising approach to stemming academic difficulties and enabling success in school (Fuchs, Fuchs, & Compton, 2013).

Many schools administer universal screeners in the form of CBM at regular intervals throughout the school year. Curriculum-based measures (CBM) are a class of assessments used to measure student progress in specific school-based skills (Deno, 1985; Deno, 2003). CBM are robust assessments of academic skills that have been standardized for difficulty within and across grade levels (Hosp, Hosp, & Dole, 2011). Measures such as oral reading fluency (ORF)

provide an estimate of students' overall reading ability by measuring the number of words a student can read per minute from a grade level appropriate text. When used as universal screeners CBM scores for individual students are compared to predetermined benchmark cut-scores (Hasbrouk & Tindal, 2006; Tindal & Nese, 2009). This comparison aids in determining if a student is at risk of failure and, therefore, requires intervention.

Currently, there are a variety of commercially and publically available CBM for use as universal screening assessments in elementary and secondary schools. The Center on Response to Intervention at American Institutes for Research (www.rti4success.org) conducts a periodic review and maintains a comprehensive catalog of CBM that includes detailed information on the psychometric properties, extent of empirical evidence, and the strength of existing evidence for each assessment. Examples include DIBELS Oral Reading Fluency (D-ORF), Reading-Curriculum Based Measure (R-CBM), Maze Reading Comprehension (Maze), Nonsense Word Fluency (NWF), and *easy*CBM<sup>™</sup> Multiple Choice Reading Comprehension (MCRC; Hintze & Silberglitt, 2005; Hosp, Hosp, & Dole, 2011; Saez, et al., 2010; Yeo, 2009).

#### **Screening in Secondary Schools**

Screening procedures at the secondary level are intended to identify students in need of additional supports and provide remediation services as early as possible. The earlier that students are identified for intervention, the more time there will be to provide interventions and potentially stem dropout. Unfortunately, the idea of *early* intervention is a bit of a misnomer in middle and high school. Once in middle and high school there is comparatively little time left in a student's K-12 career to get them back on track for graduation. In fact, as will be discussed in detail in Chapter Two, the strength of the relationship between early warning indicators and eventual dropout is so strong that it is highly unlikely that students off track in their first year of

high school will graduate without early, intensive, and ongoing intervention. Because of this time crunch, there is great urgency in identifying students that need interventions or other supplemental services as early as possible

Research demonstrating intervention effects that allow students with persistent reading deficits to make catch-up levels of growth is scarce. In 2010, Vaughn and colleagues explored the effects of a package of intensive reading interventions that incorporated decoding, vocabulary, fluency, and comprehension for students in sixth grade. Students identified for additional reading support received an average of 99.6 hours of intervention in a single school year. Though positive, effects were generally weak (median effect size .19, range .015-.19) and in most cases showed no statistically significant post-test gains over pre-test scores (Vaughn et al., 2010).

A similar study conducted by Wanzek et al. (2011) tested the effects of a year-long intervention program for middle school students with identified learning disabilities in reading. Despite daily instruction of 50 minutes in groups of 10 to 15 students, effects over the control group were small ( $\eta = .000-.054$ ) with significant effects limited only to word reading fluency.

Results of small effect sizes in the treatment of reading problems for students in middle school are not unusual. Similar results have been found in many other studies (Calhoon & Petscher, 2013; Edmonds et al., 2009; Roberts, Vaughn, Fletcher, Stuebing, & Barth, 2013; Vaughn et al., 2012)

Even when reading interventions demonstrate large positive effects it may not be enough to bring struggling readers to a level at which they no longer require additional interventions and supports. For example, a study conducted by Vaughn and colleagues on intensive reading interventions for students in grades 6-8 found that despite a treatment effect size of 1.20 most

students in the treatment group improved significantly, but did not reach grade level benchmarks of performance after three full years of intensive intervention (Vaughn et al., 2012).

These studies are presented not to be discouraging but to illustrate the difficulty in helping older students with deficits in reading achieve accelerated rates of growth and eventually reach grade level standards. In order to maximize opportunities for intervention, efficient and accurate methods of early identification of at-risk individuals are necessary.

For decades, researchers and practitioners alike have explored the connections between school failure and predictors of failure such as early literacy skills, cognitive ability, and behavior characteristics (Baydar, Brooks-Gunn, & Furstengerg, 1994). Presumably, early identification of students at risk of dropping out will enable educators to provide timely, effective intervention that will keep students in school and put them back on the path to graduation (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Pinkus, 2008). "Many students who drop out send strong distress signals for years." (Neild, Balfanz, and Herzon, 2007, p. 28). Data such as attendance, grades, state assessment results, and office discipline referrals, commonly, called Early Warning Signs (EWS), are consistently good predictors of eventual dropout (Kennelly & Monrad, 2007).

Much like CBM at the elementary level, EWS is a method of universal screening for students in need of academic and behavioral support. However, instead of administering assessments of specific skills, EWS uses existing data as an indicator of risk for failure and dropout. EWS at the upper grades are often favored over CBM due in part to the logistical considerations of data collection and time. Students in early elementary grades, specifically kindergarten and first grade, lack the archival data that is often available for older students simply because they have yet to amass such data during their limited school careers.

Furthermore, students at the early grades are also not required to take state achievement tests until third grade. Again, like CBM, EWS are purposed to detect potential achievement and behavior problems but do not typically provide information on what to do about such problems.

EWS can, however, provide insight into the nature of interventions required and help schools allocate resources toward systems and infrastructure that minimize risk factors (Pinkus, 2008). Knowing exactly which warning signs a student is exhibiting may give teachers insight into the specific services that may address the problem. For example, disaggregating early warning signs data by attendance, course failure, and office discipline referrals can help teachers determine if the best course of action is an intervention focused on developing school appropriate behaviors, providing academic support, or addressing poor attendance.

## **Screening in Middle Schools**

While the research on CBM and the use of EWS data continues to grow, there remain many unanswered questions concerning the use of CBM and EWS as they relate to universal screening in middle school. Though this has begun to change in recent years, far fewer studies exist that explore the use of CBM with students in grades 6 through 8 than at the elementary level (Denton et al., 2011). The vast majority of research in CBM as a screening tool has focused on students in elementary grades (Baker et. al, 2014; Denton et al., 2011; Yeo, 2009). Similarly, most research on EWS has focused on students in grades 9 through 12. This leaves educators working in middle schools with comparatively less guidance concerning which methods of screening are most appropriate for quickly and accurately identifying students in need of support students in grades 6 through 8. Furthermore, even amongst what is known about the reliability and predictive validity of screening measures, there are often "serious inefficiencies" (p. 265) in implementing screening procedures resulting in "unacceptably high rates of false positives (or

students who appear at risk but are not)" (Fuchs, Fuchs, & Compton, 2013, p. 266). Currently, it is not clear if either CBM or EWS are appropriately fit for students in middle school. It is also unclear which of these two methods provides the most accurate and reliable assessment of risk. In order for schools to function efficiently and effectively for all students it is necessary to explore CBM and EWS approaches to screening for students in need of additional support. The following chapter explores extant literature on three commonly used CBM assessments and four common EWS predictors, their predictive validity, and how they have been used for universal screening in middle school. This study explores two different methods (EWS and CBM) for early identification of students at-risk in reading for reading problems and how data from these two methods might be used to create a more efficient and accurate approach to screening for students in middle school.

#### **CHAPTER 2: Literature Review**

There are many popular CBM now in use across various assessment platforms. Measures such as Phoneme Segmentation Fluency (PSF), Oral Reading Fluency (ORF), and Nonsense Word Fluency (NWF), have drawn much attention from researchers given the relationship between early identification and prevention of further reading problems (Denton 2012; Goffreda, Diperna, & Pedersen, 2009; Jenkins & O'Connor, 2002). For at-risk students, early identification of reading problems may enable early intervention and therefore prevent further problems in later years (Compton, Fuchs, & Fuchs, 2013). There is now more than thirty years of research validating CBM as predictor of reading proficiency (Baker et. al, 2014; Deno, Mirkin, & Chiang, 1982; Deno, Mirkin, Chiang, & Lowry, 1980; Espin & Deno, 1995; Hintze, Keller-Margulis, & Shapiro, 2008; Helwig, Anderson, & Tindal, 2002; Hosp, Hosp, & Dole, 2011; McGlinchey & Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Yeo, 2009). Likewise, numerous studies have shown that high school dropout can be predicted with some degree of certainty using extant data from EWS (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Baydar, Brooks-Gunn, & Furstenberg, 1993; Casillas et al., 2012; Heppen & Therriault, 2008; Jerald, 2006; Pinkus, 2008). The following section outlines several of the more commonly used CBM and EWS screening measures and examines current application in middle school.

# **Universal Screening Measures in Reading**

**Oral Reading Fluency.** Oral reading fluency (ORF) is the most ubiquitous of all reading CBM (Deboy, 2013). Several online assessment platforms include measures of ORF as a key measure in their CBM assessment package for students in grades in first through eighth grade. Oral reading fluency is a general measure of grade level reading speed and accuracy. Students read grade level passages that are timed for 1-minute. ORF assessments are given in a 1-on-1

setting. Students read a passage aloud while the test administrator marks errors. The number of errors is subtracted from the total number of words read correctly to yield a score of Words Read Correctly per Minute (WRC). Students are assessed on three independent passages. The median score is then recorded as the students' score.

ORF is commonly used to assess students in first through eighth grade. The wide range of available testing levels and correlations ranging from .60 to .80 with standardized reading comprehension make ORF a popular screening assessment (Shinn & Shinn, 2002). In an analysis of ORF across grade levels, Ardoin and colleagues found a very strong relationship between ORF and other common measures of reading comprehension such as Woodcock-Johnson III (r = .76) and Iowa Test of Basic Skills-TR (r = .72) (Ardoin et al., 2004). More recently, Barth and colleagues (2012) reported that ORF correlations with external reading measures (e.g., Test of Sentence Reading Efficiency and Maze reading comprehension) were moderate to high (range r = .64 - .68) with test-retest reliability of .89 or greater.

The popularity of ORF among practitioners has drawn considerable scrutiny from researchers. When research for DIBELS, AIMSweb, and *easy*CBM are combined, there are more than 60 studies evaluating the validity, reliability, and accuracy of outcomes on ORF CBM (Goffreda & DiPerna, 2010; Yeo, 2009), nearly all of which confirm ORF as a "valid indicator of student performance on future comprehension assessments" (Deboy, 2013, p. 8).

Maze Reading Comprehension. Maze is a curriculum-based measure that assesses reading comprehension. Unlike ORF which uses reading fluency as a correlate to comprehension and overall reading achievement, Maze more directly measures students' understanding of a text. Students read a grade level appropriate passage in which every seventh word has been replaced by three words in parentheses. Students must read the passage and circle the word in parentheses

that best completes the sentence. Maze is timed at three minutes. Maze may be a groupadministered assessment. Students receive a score indicating the number of words chosen correctly and the number of words marked incorrectly (Shinn & Shinn, 2002). There are several efforts among commercial and non-profit institutions to develop Maze assessments administered digitally online. Currently, all Maze assessments available for widespread use in schools must be hand scored.

Like ORF, Maze is a common measure that is generally considered a moderately to highly reliable assessment of reading (Tolar et al., 2012). Moderate to high correlations have been reported with established measures of reading comprehension such as Woodcock-Johnson Passage Comprehension Subtest (r = .56) and Iowa Test of Basic Skills Reading Comprehension (r = .70; Ardoin et al., 2004; Parker, Hasbrouk, & Tindal, 1992). Espin and colleagues confirmed these results for students in middle school reporting validity coefficients r = .70 or greater in predicting scaled scores on statewide assessments of reading achievement (Espin, Wallace, Lembke, Campbell & Long, 2010).

Maze provides a combined assessment of silent reading fluency and reading comprehension (Fuchs & Fuchs, 1992; Tolar, et al., 2012). The measurement of comprehension as a part of CBM screening assessments is appealing over oral reading fluency, particularly to teachers in upper elementary and middle school. As students move into upper grades, measuring fluency alone may be undesirable as the focus of instruction shifts from learning to read to reading to learn (Ardoin et al., 2004; Espin et al., 2010; Tolar, et al., 2012).

Prior research on Maze has focused largely on students at the elementary level. However, the evidence base has expanded in recent years to include students in middle school (Tolar, et al., 2012). Most notably, a 2012, investigation by Tolar and colleagues explored the psychometric

properties of Maze for students in grades 6 through 8 including test form variability, and predictive validity. Investigators reported predictive validity coefficients consistent with previous research conducted at the elementary grades ranging from .54 to .73. Results for test-reliability were also consistent with previous research with coefficients ranging from .73 to .91 (Tolar, et al., 2012).

*easy*CBM Multiple Choice Reading Comprehension. *easy*CBM Multiple-Choice Reading Comprehension (MCRC) is an untimed assessment that measures reading comprehension on grade level appropriate text passages. After reading a passage, students answer factual, inferential, and analytical questions. For students in third through eighth grade MCRC contains a total of 20 questions. Text passages are made available while answering questions. Students may refer to the text as often as needed throughout the assessment. MCRC is designed to be group administered and may be given to an unlimited number of students simultaneously via online assessment modules or hardcopy formats (Riverside, 2012). Assessments completed using the online module are scored instantly. Hardcopy assessments must be scored by hand. Scores are reported as the total number of questions answered correctly.

Unlike other forms of curriculum-based measurement such as ORF and Maze, that assess basic skills correlating highly with grade level content skills and knowledge, MCRC directly measures grade level reading comprehension skills. MCRC is the most common format of reading comprehension assessment (Andreassen & Braten, 2010), similar to many other measures of reading comprehension including many statewide assessments of reading throughout the country (Deboy, 2013). Reports on MCRC show technical adequacy in alternate form reliability, Rasch item analysis, and split half reliability (range .12-.63; Irvin, Alonzo, Lai, Park, & Tindal, 2012).

#### **Early Warning Signs**

Early Warning Signs (EWS) is the common name for a variety of data sources that have been found to correlate highly with high school dropout (Kennelly & Monrad, 2007). The National High School Center, Alliance for Excellent Education, National Middle School Association, and American Institute for Research actively promote the use EWS as a mechanism for identifying students for interventions and supplemental services to prevent dropout. Funded by the U. S. Department of Education's Office of Elementary and Secondary Education and Office of Special Education Programs, the National High School Center (NHSC; BetterHighSchools.org) produces free software (i.e. Early Warning Systems) designed to assist districts in compiling EWS data in order to identify students at risk of dropout. Using course performance, attendance, and credit accrual, school personnel can use EWS to assess students' overall risk status and identify students with multiple at-risk factors (Heppen & Theriault, 2008).

Though the majority of EWS attention has been devoted specifically to the high school level, identifying students in need of support as as late as high school may not provide sufficient time to provide intervention services and get students on track for graduation. To give students and teachers more opportunities for remediation, many have recognized the need for identification of students at-risk before high school (Belfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Casillas, et al., 2012; Neild, Balfanz, & Herzog, 2007; Pinkus, 2008).

In 2007, Neild, Balfanz, and Herzog examined commonly collected data from school districts in hopes of finding reliable predictors of which students would eventually drop out of school. Researchers examined grades, test scores, attendance, behavior reports and special education status for a cohort of students from sixth grade through graduation. The team determined that over half of eventual dropouts were identifiable as middle school students. In

one investigation researchers found that, for students in sixth grade, anyone that exhibited one or more of any of the following markers; (a) failure in mathematics, (b) failure in English, (c) attendance rate below 80%, or (d) "unsatisfactory mark in behavior" for at least one class, had a 75% chance of dropping out. When researchers examined data from students in eighth grade, one course failure in English or mathematics or below 80% attendance once again signaled students with a 75% chance of dropout (Neild, Balfanz, & Herzog, 2007).

In a follow-up study two years later Balfanz and colleagues extended the exploration of early warning signs by replicating the research in five additional school districts. In all replications the predictors used as early warning signs for students in eighth grade from the previous study now identified students with a 25% or lower chance of graduation. This follow-up study also added two points of consideration. First, students exhibiting warning signs in sixth grade and eventually dropping out of school did so typically during their junior year of high school. That means that students that showed warning signs in sixth grade remained in school for five additional years. This indicates that schools have a substantial window of opportunity to intervene and potentially prevent eventual dropout. Second, researchers explored the use of course failures in the form of Fs (grades below 60%) as well as near failures or D's (grades between 60-69.99%). Grades of F are generally considered as failure for which students receive no credit accrual. Passing courses is, "key to earning the required credits to graduate" (p. 5, Belfanz, 2009). Grades of D are just above failure and thus considered passing. However, grades of D were predictive of Fs, and therefore a potential signal that predicts dropout before failures actually develop. The inclusion of both Ds and Fs produced a model that more confidently captured all potential students at risk. However the model also produced a higher proportion of students flagged for potential dropout than eventually dropped out. Many of these students would

have not been incorrectly identified using F's alone. The authors contend that a higher potential for false positives risks is worth the additional security gained by reaching more students actually at-risk for dropout (Belfanz, 2009).

Attendance. Attendance is an early warning sign with direct connection to academic proficiency and successful graduation. Absences present a physical barrier to learning in the school environment. If students are not present for instruction, their access to content is limited to what they can manage independently. Though this may be less of a concern in environments where online content is the primary mechanism of instruction.

Attaining proficiency in course content can be challenging, even when students are in attendance. Absence from school compounds existing learning difficulties or often causes learning difficulties due to lack of instruction. Allensworth and Easton cite absences as a strong predictor of failure for students in ninth grade, noting a strong negative relationship  $R^2 = -.51$  between absences and grade point average (2005). Even moderate rates of absences (5-10 days per semester) are associated with increased rates of dropout (Heppen & Theriault, 2008).

Neild and Belfanz (2006) found the number of absences accrued in the first 30 days of high school was one of the strongest predictors of eventual dropout when compared to other risk factors such as gender, race, age, and scores on standardized tests. Similarly, in a long-term study with Philadelphia Public Schools, researchers found that among students that eventually dropped out of school, 85% exhibited patterns of absences that developed in sixth grade and increased throughout the remainder of their careers in middle school (Neild, Belfanz, & Herzog, 2007).

**Behavior.** Like attendance, patterns of inappropriate behavior can be both a direct problem in and of themselves, and a factor that exacerbates existing learning difficulties. Behaviors that detract from learning limit a student's ability to learn by focusing attention

elsewhere and preventing engagement in instruction. Inappropriate behavior can impede the learning of other students and the teachers' ability to teach effectively. Comorbidity of academic and behavior problems is well documented, spanning literature in dropout prevention, special education, behavior management, health, neuroscience, and others (Barry, Lyman, & Klinger, 2002; Hinshaw, 1992; McIntosh, Goodman, & Bohannon, 2010; Nigg, Hinshaw, Crarte, & Treuting, 1998; Reinke, Herman, Petras, & Ialongo, 2008; Sullivan, Childs, & O'Connell, 2010; Tobin & Sugai, 2009).

Given this connection between behavior and academic performance, schools can use behavior as a predictor of academic risk. This is typically done using data from school suspensions (Jerald, 2006). In an exploration of dropout predictors using grade point average, race, socioeconomic status, and suspensions, researchers showed that a single suspension increased the likelihood that a student would drop out by 77.5% (Suh & Suh, 2007).

However, data from suspensions is only one way in which behavior data can serve as a predictor of academic risk. Office Discipline Referrals (ODRs) can also be used to identify potential academic problems. ODRs refers to the documentation of major and minor behavioral infractions such as fighting, harassment, and noncompliance. ODRs are a source of data that can be useful for to identifying students with behavior difficulties and to inform the problem solving process (Sugai, et al., 2000).

In 2009, Tobin and Sugai conducted a study that supports the use of discipline referrals as a method for predicting future academic and behavior problems. Researchers analyzed longitudinal data from the school records of 526 students to predict negative high school outcomes. Researchers created a model using data from sixth grade records to predict whether or not students would be on track for high school graduation in ninth grade. The model included

discipline referrals for fighting, harassment, nonviolent misbehavior, grade point average, contact with the juvenile justice system, and out-of-school suspensions. When controlled for variance related to gender, analysis showed that referrals were a statistically significant predictors ( $r^2 = .65 - .91$ , p = <.001) of whether or not students were on track academically for graduation, as early as ninth grade (Tobin & Sugai, 2009).

In addition to the strong correlation between academics and behavior, ODRs have other properties that make these data useful for screening purposes. First, ODR is not synonymous with punishment. Tracking ODRs differs from suspensions in that ODRs are logged based on the occurrence of an event warranting documentation as opposed to an event that warrants temporary exclusion from the school environment. Students can also accumulate ODRs for minor infractions, meaning that schools can use ODRs to identify students that exhibit patterns of behavior not severe enough for suspension.

**Course Failure.** Failing grades are another key indicator of high school dropout. At the secondary level course grades are the primary factor behind credit accrual, grade point average, and career and college readiness. Without passing grades, it is unlikely that students will accrue sufficient credits for graduation and will not reach acceptable levels of academic proficiency required for post-secondary education or the workplace (Casillas et al., 2012). When compared to all other students, students that exhibit even a single course failure in ninth grade show increased rates of future course failure, decreasing test scores and increases in behavior problems (Cohen & Smerdon, 2009).

Even in middle school, course grades remain an important variable in monitoring students' long term and short-term academic progress toward graduation. The Consortium on Chicago School Research reports that a single course failure is highly predictive of eventual

dropout, noting that 83% of all students with no failed courses their freshman year will go on to graduate (Allensworth & Easton, 2005). When data on course grades are examined with other risk predictors, the reliability of correct prediction increases substantially. Casillas et al. found that course failure in eighth grade mathematics and English along with attendance lower than 80% identifies almost 80% of all future high school dropouts (Casillas et al., 2012).

Fortunately, schools that target students at risk for failure and provide additional academic supports are generally successful in lowering high school dropout (Cohen & Smerdon, 2009; Neild, Belfanz, & Herzog, 2007, Belfanz, 2009). When used as a source of data for screening students at risk of failure in reading, course failure is an important factor in identifying students in need of support and preventing high school dropout.

Statewide Reading Assessments. Since 2002, schools are required to meet federal school assessment and accountability standards under No Child Left Behind (NCLB) PL 107-115, or file for a federal wavier with a state and federally approved school improvement plan. States must assess students in third through eighth grade in math and reading every year. School-wide assessment results must be publically available, with student level results made freely available to parents. Assessment results are then used to determine to what extent schools have been successful in teaching grade level content standards to all students. As the primary driver of school accountability, these results weigh heavily into school's overall measure of effectiveness. Schools that do not reach Adequate Yearly Progress (AYP) may incur sanctions that include loss of state funding, state imposed restructuring or reconfiguration, and building closure (Hunley, Davies, & Miller, 2013; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2004; Yeo, 2009).

Research exploring the reliability and validity of CBM has a well-established history of using state assessment data as the dependent variable. (Baker et. al, 2014; Deboy, 2013; Denton

et al., 2011; Tindal, Nese, & Alonzo; Yeo, 2009; Wiley & Deno, 2005). As of 2014 there were more than 30 studies in oral reading fluency alone that used state administered reading assessments from 15 different states as the dependent variable in analysis (Kilgus, Methe, Maggin, & Tomasula, 2014; Yeo, 2009).

However, what is unique about statewide reading achievement tests is that they can be considered both a measure of outcomes and a measure of prediction. Statewide achievement tests are often considered *high stakes* given the potential sanctions levied on schools that fail to reach achievement goals. Many school districts now include state-wide assessment results as part of teacher evaluation process, identification of students for intervention, and even as replacements for scientifically validated reading diagnostics (Casillas et al, 2012; Smartt & Reschly, 2007). It is also common to have annual IEP goals for students with disabilities tied to results on statewide tests (Katsiyannis, Zhang, Ryan, & Jones, 2007). In this sense, state mandated reading achievement tests are one of the outcomes on which schools are trying to effect positive change.

Conversely, scores on state achievement tests can be a predictor of future academic performance. In fact, the psychometric procedures used in determining benchmark cut scores (i.e. logistic regression, signal detection theory, and equipercentile matching) attempt to define with some certainty how students are likely to score on future tests given their current performance. This essentially means cut scores are set such that current performance relative to the cut score predicts future performance. If, by design, state test scores predict future test scores, schools may be able to use scores on state assessments to serve as a screening method for identifying students that are not likely to reach grade level standards without some type of intervention or supplemental service.

Therefore, prior performance on state tests may serve the same function as CBM and EWS. Reed, Wexler, and Vaughn argue that state assessments alone may be an efficient screening tool for predicting which students are currently at risk for not reaching grade level benchmark and eventual dropout (2012). Since state assessments are mandated, schools are already required to allocate resources to comply with federal and state accountability law. Using state test data as a screening measure might then be a very efficient use of data from a test students are already required to take. State assessments of reading are administered annually and generally include criterion-referenced benchmarks of performance that may approximate the results of a dedicated screening assessment. Many state assessments also report a confidence interval or range of performance that can be considered in order to widen the potential pool of students that may need intervention (Reed, Wexler, & Vaughn, 2012). Prior research also indicates that past performance on state assessments is a good predictor of future performance on state assessments. In an analysis of ten different measures of reading, Denton and colleagues (2011) reported that of 10 measures, the best predictor of 2007 Texas reading comprehension accountability test (TAKS) scores was 2006 TAKS scores (AUC = .82). Additionally, 44% of variance on the TAKS was explained by TAKS scores from the previous year.

#### **Recent Research in Middle School Screening**

By comparison, far fewer studies have examined the use of CBM and other methods for risk identification (i.e. EWS) at the middle school level than exist at the elementary and high school levels. While CBM for early grades were developed with specific attention paid to key skills in early literacy they may not be appropriate for adolescent learners. Moving from learning basic skills to more complex grade level content, middle schools may require CBM that address this shift. That is not to say that reading CBM developed for use at the elementary level are

inherently ill-fit for use at the secondary level (Baker et. al, 2014; Yeo, 2009). On the contrary, skills that continue to develop throughout a student's scholastic career such as oral reading fluency and basic reading comprehension may in fact be useful measures for universal screening (Helwig, Anderson, & Tindal, 2002). Despite a popular belief in the diminishing relationship between fluency and grade level reading comprehension as students age (Baker et. al, 2014). In 2009, Yeo conducted a meta-analysis of oral reading fluency CBM and found no such relationship. At the time however, there were only two published studies that included students beyond fifth grade. The overwhelming majority of data represented students in grades three and four. A more recent study by Denton et al. (2011), found correlations between oral reading fluency and high stakes tests (r = .69-.76) for students in grades 6 through 8 were comparable to correlations typically found at the elementary level. Fortunately, research is now expanding to include more data from secondary grades and address these questions more thoroughly.

Baker and colleagues (2014) recently explored the diagnostic efficiency and criterion validity of CBM in measures of oral reading fluency, word reading accuracy, and multiple choice reading comprehension. Researchers collected data from 2,943 students in grades 7 and 8. Using state assessment results in reading as the outcome measure, researchers determined that oral reading fluency and reading comprehension combined, led to more accurate diagnosis of reading failure than either measure alone. The combination of ORF and reading comprehension also explained the greatest amount of variance in outcomes compared to all other tested combinations of predictor variables. Word reading accuracy provided little additional value to the predictive model.

Similarly, Stevenson (in press) examined the accuracy and reliability of oral reading fluency, maze reading comprehension, and *easy*CBM multiple choice reading comprehension

(MCRC) in predicting outcomes on state assessments. Using data from 494 students in grades 7 and 8, researchers examined which combination of measures most accurately predicted outcomes on state reading tests in Michigan. Results indicate that for students in seventh grade a combination of *easy*CBM MCRC and oral reading fluency provided the highest significant level of classification accuracy (77%) and accounted for greater variance (53.2%) than any other linear combination of measures tested. Results were similar for students in eighth grade with *easy*CBM MCRC and oral reading together producing a classification accuracy of 82.3% with variance explained at 60.7%. Results indicate the combination of MCRC and oral reading fluency provided the best overall predictor of reading proficiency among the combinations tested.

### **Research Questions**

It is clear from the extant literature that there are a number of predictors from both *new* assessment data (CBM) and extant records (EWS) in middle school that correlate with latter success and/or failure in school. What is unclear, however is which of these approaches, CBM or EWS (and the individual predictors therein) are the most suitable for use in middle school. To determine the most effective screening strategy there are many variables to consider including cost, time devoted to collecting the data, the personnel and training required, logistical considerations, and the accuracy of screening procedures in correctly identifying which students do or do not require addition supports to reach academic standards and eventually graduate. The following investigation explores this question of accuracy in screening, specifically targeting reading proficiency and outcomes on high-stakes assessments. The current study is a direct comparison of CBM and EWS approaches for identification of students at risk of failure in reading for grades seven and eight.

The primary research questions for this investigation are:

- 1. How do curriculum-based measures in reading compare to early warning signs data in predicting outcomes on high stakes tests of reading at the middle school level?
- 2. What is the valued added in combining curriculum-based measures (CBM) in oral reading fluency, Maze reading comprehension, and *easy*CBM multiple choice reading comprehension as predictors of outcomes on high stakes reading assessments at the middle school level?
- 3. What is the valued added in combining Early Warning Signs (EWS) data in attendance, office discipline referrals, failing grades, and prior performance on statewide assessments function as predictors of outcomes on high stakes reading assessments at the middle school level?

### **CHAPTER 3: Methods**

# **Participants and Setting**

All data for this investigation were collected from a suburban middle school in the Midwest. The school serves 434 students in grades seven (n = 197) and eight (n = 237). Enrollment includes 48.8% male, 51.2% female. Enrollment by race/ethnicity is 4.7% of students identified as Asian, 24.1% African-American, 36.8% Caucasian, 19.6% Hispanic, and 14.6% identified as two or more races. Less than 1% of students were categorized as other or unreported. Additional demographic reporting includes 64.3% of students eligible for free or reduced lunch (FRL), 2.2% of students qualified as English Language Learners, and 15.7% of students enrolled in special education.

The site for this investigation was chosen based on the capacity of the school and district to provide the appropriate data for the questions of interests. The researcher for this investigation is formerly the Data Coach for the participating school district. Though the researcher participated in administration of many of the assessments included in the dataset, the researcher was not directly involved in compiling the dataset used for the current investigation.

## **Data Collection**

Data for this study were gathered exclusively from archival sources. School staff supplied the researcher with the dataset electronically. Prior to delivering the dataset to researchers, school staff removed all students' identifying information including names, birth dates, local school identification numbers, and state identification codes. All assessment scores and demographic information included in this study were collected as part of the participating school's regular operating procedures. Assessment scores were collected as part of the typical battery of assessments given to all students in fall of each academic year. Since all data were

collected as part of regular educational practice and did not contain identifying information, the study posed no risk for students and did not fit the definition of human subjects research. Therefore, the current study was determined exempt from review by the Institutional Review Board.

Data supplied to the researcher included demographic information for grade level, gender, race, eligibility for free or reduced price lunch, disability status, disability classification, and status as English Language Learners. District staff also provided data for attendance, office discipline referrals, and course failures for the 2013-2014 school year. Assessment data were provided for three universal screening assessments administered in September 2013 including oral reading fluency, Maze reading comprehension, and *easy*CBM Multiple Choice Reading Comprehension. The participating school routine administers oral reading fluency, Maze, and easyCBM assessments three times each year. Statewide assessment data for the Michigan Education Assessment Program (MEAP) reading subtest were provided for fall 2012 and fall 2013 testing periods. MEAP is administered annually to all public and charter schools in the state of Michigan. Local district staff for each school does administration of the MEAP in each school. Data from the MEAP reading subtest included raw score, scaled score, and proficiency level. Data from ACT Explore<sup>®</sup> reading subtest included raw score and scaled score. ACT Explore<sup>®</sup> is administered to students in the Spring of the eighth grade year. ACT Explore<sup>®</sup> is included as part the districts assessment of career and college readiness. The sequence of career and college readiness includes ACT Explore<sup>®</sup> (eighth grade), ACT Plan<sup>®</sup> (tenth grade), and ACT<sup>®</sup> (eleventh grade). Detailed purpose and rationale for the inclusion of each of these assessments is included below.

All students enrolled during the assessment periods participated in all assessments. Missing scores in the dataset were due to changes in school enrollment for individual students. Any students with a missing score in the dataset became enrolled in the school after the conclusion of benchmark testing or was provided an alternative statewide reading assessment in compliance with state testing guidelines and the students' Individualized Education Plan (IEP).

The participating school provides formal training to all staff involved in direct administration of CBM and state assessments. All teachers participate in administration of state assessments. All teachers also participate in administration and scoring of CBM of at least one type of assessment. Not all teachers participate in administration of all CBM assessments, though everyone is required to participate in administration of at least one measure. The district also direct assistance to teachers during the test periods as part of the training protocols. Though the participating school provided training for each individual responsible for administration of CBM and state assessments, at the time of this investigation they did not systematically collect fidelity of implementation data for adherence to testing procedures. Therefore it was not possible to determine fidelity of testing procedures.

## **Reading Screening Measures**

The participating school conducts universal screening in reading and math three times each year in September, January, and May. Each screening period (also known as benchmark testing period) includes a two-week window of opportunity for initial testing and an additional week for make-up testing. All reading screening assessment data for the current investigation were from the fall screening period. For the purposes of this investigation, the fall administration of universal screening measures in reading was the period that immediately preceded

administration of state reading tests. There was approximately three weeks separating the end of the benchmark testing period and start of administration of MEAP.

*easy*CBM Multiple-Choice Reading Comprehension. All assessments of Multiple Choice Reading Comprehension were conducted using *easy*CBM MCRC assessment (Riverside, 2012). Group administration of MCRC was done electronically using iPads. Each teacher that participated in the administration of MCRC received training in testing protocols and procedures from the district level *easy*CBM system administrator. Prior to each benchmark testing period, each participating teacher received a brief refresher training in the administration procedures. All students took the MCRC during the predetermined benchmark assessment period for fall 2013. *easy*CBM MCRC consists of 20 multiple choice reading comprehension questions. Each question was worth one point. Scores for MCRC were recorded in the dataset as raw scores.

**Reading-Curriculum Based Measure.** Reading-Curriculum Based Measure (R-CBM) is a measure of oral reading fluency. All R-CBM assessments were administered using the *AIMS*web browser based scoring system. R-CBM is given in a 1-on-1 setting. The participating school uses a team-based approach to administering benchmark assessments. A group of five teachers, paraprofessionals, and instructional coaches conducted all administrations of R-CBM benchmark testing. All school staff participating in benchmark assessments receive ongoing training in the administration and scoring of reading measures within the *AIMS*web system. During benchmark administrations of R-CBM, each student reads three grade level passages timed at 1-minute each. The median score was then recorded. Median scores were reported to researcher in the form of, (a) number of words read correctly per minute, (b) number of errors, and (c) percentage of words read correctly.

Maze Reading Comprehension. Maze was group administered in hardcopy format during language arts classes. All language arts teachers previously received training in the administration and scoring of Maze assessments. All language arts teachers also received a brief refresher in training protocols no more than one week prior to the start of the benchmark testing period. All Maze assessments were scored by the same team responsible for the administration of R-CBM assessments. Data were provided to researchers for the total words marked correctly, total errors, and percentage accuracy.

#### **Early Warning Signs**

Attendance. Attendance data for this investigation were gathered directly from the participating school's student information system. Absence data from the 2012-2013 school were provided for all students. Per school policy, all teachers are required to record student absences and late arrivals at the beginning of each of 7 periods throughout the day. According to school policy, late arrivals less than ten minutes late are coded as "T," for tardy. Late arrivals greater than 10 minutes late but less than 20 minutes late are coded as "Y" indicating a severe tardy. Any late arrival greater than 20 minutes is coded as an absence. Absence codes are assigned to each absence according to the circumstances behind the absence. Attendance codes included A = Absent-Unexcused, I = In-School Suspension, S = Out-of-School Suspension, U = Unexcused Absence, and V = School activity. For the purposes of this investigation all absence codes were collapsed to a single code. All absences codes were aggregated to yield the total number of absence codes accumulated in the most recent school year.

Absences of any type indicate that the student was not present for instruction. Absence from instruction is a significant predictor of academic failure (Balfanz, 2005).
**Office Discipline Referrals.** Office discipline referrals (ODR) is the term associated with all major and minor disciplinary incidents that occur within the school environment. School staff track ODRs for all students using the School-Wide Information System (SWIS) digital online behavior tracking system developed at Education and Community Supports (ECS) at the University of Oregon (pbisapps.org). At the participating school, all instances of lunch detention, after-school detention, Saturday school, in-school suspensions, out-of-school suspensions, expulsions, and referrals to the school's time-out room are logged as ODRs. Data on location, date, time of day, referring staff member, behavior warranting the referral, and likely function of given behavior are logged with each ODR. To preserve confidentiality, only the total number of ODRs for each student were provided to the researcher.

**Failing Grades.** Data for course failures were collected from the school's electronic student information system. All final course grades from the prior school year were provided to the researcher as letter grades (A, B, C, D, & F). The total number of course failures was calculated for each student. Though the participating school considered a grade of D as passing, the inclusion of D as a failure in screening procedures decreases the likelihood of misidentifying risk status for students that may be near failure but are technically passing courses. Inclusion of D grades as failures is also consistent with research and policy recommendations of early warning signs. (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007).

#### **Reading Outcome Measures**

**Michigan Education Assessment Program.** The Michigan Education Assessment Program (MEAP) was the statewide system of assessments for academic progress for the State of Michigan. MEAP annually assessed grade level content in reading, writing, math, science, and social studies in the fall of each academic year for third through ninth grades. Fall administration

of MEAP occurred approximately two weeks after the close of the benchmark screening period for universal screening.

The reading subtest of the MEAP assessment was selected as the primary dependent variable in this investigation for several. First, the cut scores used to determine the difference between proficient and not proficient were revised in 2011 to become non-arbitrary points. Cut scores are backward mapped for each grade level beginning at a cut point for students at eleventh grade with a 50% probability of earning B or better in their first college course. All other cut scores are determined by a 50% likelihood of reaching proficiency on the following year's MEAP test (State of Michigan). Second, as the assessment used for school accountability under federal requirements MEAP is indeed a high stakes test. As the statewide reading assessment, MEAP is aligned with the state adopted standards for reading at each grade level. Fourth, all students in third through ninth grade were required to take MEAP unless the student had an active IEP and the students' "Individualized Education Plan (IEP) teams determined it was not appropriate for them to participate in the state's general education assessments" (p. H-3, Michigan Department of Education, Statewide Assessment Selection, [MDE] 2012). One student in the dataset was marked by the school as eligible for an alternative assessment and did not take MEAP Reading.

School staff indicated that for all students with disabilities that qualified to take MEAP, all testing accommodations were provided to students in accordance with each students' IEP requirements. Data on which students received which accommodations were not provided to the researcher.

Results for MEAP Reading were provided to the researcher in the form of scaled scores and proficiency levels. Proficiency levels were determined by the Michigan Department of

Education, Bureau of Assessment and Accountability (BAA). Scale scores for each student were organized according to predetermined performance criteria. Performance levels include; 1 = advanced, 2 = proficient, 3 = partially proficient, and 4 = not proficient. Only students scoring advanced and proficient are counted as proficient under Michigan State Accountability Criteria.

Technical documentation for MEAP report statewide classification accuracy in reading of 73.9% for seventh grade and 74.7% for eighth grade. Empirical Item Response Theory (IRT) statistics for reliability range from  $\alpha$  = .80-.82 for seventh and eight grade (Michigan Department of Education, 2012).

For the 2014-2015 school year, MEAP was discontinued by the Michigan Department of Education. As of spring 2015 the statewide assessment of reading achievement and accountability is the Michigan Student Test of Educational Progress (M-STEP). M-STEP is administered in the spring of each school year.

ACT Explore<sup>®</sup> Reading. ACT Explore<sup>®</sup> is a nationally administered assessment of academic achievement. ACT Explore<sup>®</sup> is the middle school companion of the ACT<sup>®</sup>, <sup>a commonly used college entrance exam. ACT Explore<sup>®</sup> was selected as a secondary outcome measure for this investigation for several reasons. First, the purpose of ACT<sup>®</sup> assessments is to assess students' readiness for college. Scaled scores for ACT Explore<sup>®</sup> are aligned to the outcomes of ACT Plan given in tenth grade and the regular ACT<sup>®</sup> taken for college entrance requirements in the eleventh grade. That means that a score of 16 on the ACT Explore<sup>®</sup> is equivalent to a score of 16 on the college entrance ACT<sup>®</sup>. This is advantageous in that students' current state of college readiness in eighth grade can be assessed in a way that is directly comparable to the test students take for their actual college readiness salient as early as eighth grade. Inclusion of ACT</sup>

Explore<sup>®</sup> as a dependent variable in this investigation is therefore helpful in evaluating the EWS and CBM prediction as they relate specifically to college and career readiness. Second, MEAP is a test of proficiency on state adopted skills and standards. ACT Explore<sup>®</sup> is a nationally administered test that is not directly tied to a single state's adoption of standards. Instead, ACT Explore<sup>®</sup> assesses literacy and math skills more broadly since it must be configured to cover standards across multiple states. This, analysis of ACT Explore<sup>®</sup> provides a unique opportunity to readily compare the results of the prediction models under investigation with both state level and national level data.

ACT Explore<sup>®</sup> is administered to students in the spring of eighth grade. Scores for ACT Explore<sup>®</sup> were reported as raw scores and scaled scores. The proficiency cut points for career and college readiness used for scaled scores on ACT Explore<sup>®</sup> were determined by ACT<sup>®</sup> (ACT, 2013). As of June 2014 ACT Explore<sup>®</sup> is no longer available. ACT Explore<sup>®</sup> has been replaced by ACT Aspire<sup>TM</sup> beginning spring 2015.

# **Data Analysis**

**Data Screening.** Prior to analysis, all assessment data were examined for outliers, normality in distribution, skewness and kurtosis for each dependent measure. All statistical analyses were conducted independently for each grade level, outcome measure, and method of analysis. The dependent variables were MEAP Reading and ACT Explore<sup>®</sup>. In all analyses, cases with missing data were excluded pairwise. Analyses were conducted using hierarchical logistic regression and hierarchical linear regression.

**Binary Analysis.** The chief function of any universal screening assessment is to determine if a student is at-risk or not at-risk for academic problems. Therefore it is necessary to analyze how accurately screening procedures correctly predict categorization of student as at-risk

or not at-risk for not achieving grade level standards of performance. The current study uses EWS and CBM models to predict risk categorization in reading. Each assessment of risk predicted by each model is compared to the actual outcomes of students' reading proficiency on high stakes reading tests. Since the primary outcome of consideration is the accuracy in predicting proficiency on high stakes assessments, the outcome of greatest interest is dichotomous (proficient or not proficient). For hypothesis testing of categorical and continuous predictors on dichotomous outcomes logistic regression is the most appropriate method of analysis (Peng, Lee, & Ingersoll, 2001). Scores for each of the dependent variables (MEAP reading and ACT Explore reading) were collapsed into binary outcomes. MEAP proficiency levels (1) advanced and (2) proficient were combined into a single category of (1) proficient. MEAP proficiency levels (3) partially proficient and (4) not proficient were combined into the category of (0) not proficient. ACT Explore<sup>®</sup> scaled scores were categorized based on the college readiness benchmark cut scores provided by ACT® (ACT, 2013). Scaled scores greater than or equal to 16 were recoded (1) on-track for college readiness Scaled scores less than 16 were recoded as (0) not on-track for college readiness.

Hierarchical Logistic regression was used to assess the accuracy and value added of CBM and EWS variables, individual and in combination, in predicting outcomes on MEAP and ACT Explore. The enter method was used to add predictor variables separately for each block of the prediction models. In order to control for variance due to demographic variables, ethnicity, gender, and students qualifying for free or reduced lunch (FRL), were entered in block one of each analysis. The inclusion of demographic variables in subsequent blocks of the regression models was contingent upon statistical significance. Significant ( $\beta$ -value, p < .05) demographic

variables were retained in each subsequent block. Non-significant variables ( $\beta$ -value, p > .05) were removed from the model.

Order for block entry of CBM variables was based on the results of prior CBM research for students in middle school (Baker et al, 2014; Stevenson, in press). Variables were entered in order of the strength of predictive relationship with the outcome measures, with the strongest predictors entered first and the weakest predictors entered last. Block two includes *easy*CBM MCRC. Block three includes R-CBM. Block four includes Maze reading comprehension. The CBM prediction model was run independently for MEAP and ACT Explore at each grade level.

A separate prediction model was created and tested for EWS variables using logistic regression. Block entry of variables for early warning signs data is based on variance explained, likelihood of correctly predicting school dropout from extant data (Belfanz, 2009; Belfanz, Herzog, Mac Iver, 2007; Heppen, Therriault, 2008; Jerald, 2006; Pinkus, 2008), and the use of past scores on state assessments to predict future scores on state tests (Casillas et al., 2012; Denton et al., 2011; 2012; Reed, Wexler, & Vaughn, 2009).

Demographic variables were entered into the first block of the prediction model as described above. Block two includes the most recent MEAP reading scale score (MEAP-Y1). Block three includes the total number office discipline referrals for the prior school year (Tobin & Sugai, 1999). Block four includes attendance in the form of the total number of period absences from the prior school year. Block five includes the total number of course failures. Course failures include grades of both D and F, commensurate with the findings of Belfanz and colleagues (2007). The EWS prediction model was run independently for MEAP and ACT Explore at each grade level.

Linear Analysis. Hierarchical linear regression was also used to examine the relationship between combinations of EWS and CBM predictor variables and outcomes on MEAP reading and ACT Explore Reading. Scaled scores for MEAP an ACT Explore were used as the dependent measures. Analyses were conducted for combinations of EWS predictors and combinations of CBM independently. Block entries of variables for each hierarchical linear regression model were entered in the same order as in the logistic models described above. Linear analyses were run separately for each grade level.

### **CHAPTER 4: Results**

# **Data Screening**

A summary of descriptive statistics and tests for normality are included in Table 1. Data for ACT Explore<sup>®</sup>, MEAP Reading, R-CBM, easyCBM, MCRC, and Maze were within ±1.5 for skewness and kurtosis (Tabachnick & Fidell, 2013). Examination of Q-Q plots and histograms indicate that the data were normally distributed, therefore no data transformations were performed (George & Mallery, 2010; Ghasemi & Zahediasl, 2012). However, seventh and eighth grade scores for easyCBM were significant on the Shapiro-Wilk test indicating a potential nonnormal distribution (eighth grade = .951, p < .001; seventh grade = .928, p < .001). MEAP scores for both grades also showed significant test statistics on Shapiro-Wilk (eighth grade = .955, p < .001; seventh = .986, p = .020). Some indication of non-normality was anticipated given that the school-wide percent proficient for MEAP reading was below the state average for both seventh (-11.7%) and eighth (-6.5%) grade. However, as normality in distribution is not an assumption that must be met for logistic regression, no transformations of the data were necessary (Peng, Lee, & Ingersoll, 2002). All other measures of reading were within recommended tolerances for normality; thus it is unlikely that the potential for non-normal distribution of data would compromise the following statistical analysis.

## **Demographic Variables**

In each of the screening models under investigation the demographic variables of gender, socio-economic status (SES; as measured by students qualifying under federal guidelines for Free or Reduce Lunch), and ethnicity were included in the first entry of variables into each model. Achievement gaps among these subgroups are well documented in the literature (Hernandez, 2011; O'Connor & Fernandez, 2006; O'Connor, Hill, & Robinson, 2009). In

particular, students from low-SES households, Hispanic students, and African-American students typically are disproportionately identified at-risk and have fewer students identified as gifted and talented (O'Connor & Fernandez, 2006; O'Connor, Hill, & Robinson, 2009; Hernandez, 2011). According to Hernandez (2011) 22% of all students that live in poverty at any time throughout their childhood do not graduate from high school. Data from the National Assessment of Educational Progress (NAEP) from 1992 to 2009 showed an average gap of 25 point gap in reading achievement between White and Hispanic students in eighth grade (Hemphill, Vanneman, & Rahman, 2011). Results are similar for the gap in reading achievement between White and Black students. NAEP scores from 1992 to 2007 showed a 7-point deference between White and Black students with no statistically significant change between 1998 and 2007 (Vanneman, Hamilton, Anderson, & Rahman, 2009).

While this investigation is not specifically concerned with the associations between demographic subgroups and student achievement, it would be a mistake to ignore variables that are known to be associated with increased risk for failure and dropout. In order to appropriately evaluate the prediction models of interest, demographic variables were included in each prediction model in order to control for these effects. In subsequent entries of variables for each model, demographic variables that returned non-significant statistics for variance explained were removed from each model beginning in block two.

#### **Binary Analysis**

**Early Warning Signs.** Block entry of EWS variables in the prediction model were entered in the following order: demographic variables, prior scores on state reading assessments (MEAP-Y1), ODRs, attendance, and course failure. Model fit was assessed for each block entry of variables. Test statistics for model fit and variance explained for outcomes on MEAP reading

are summarized in Table 2 (seventh grade) and Table 3 (eighth grade). Chi-squared and Hosmer-Lemeshow (H-L) tests were preformed for goodness of fit.

For seventh grade EWS predicting MEAP, H-L tests for all five blocks were nonsignificant indicating that the models were not poorly fit to the data (see Table 2). Omnibus tests showed significance in block one (FRL),  $\chi^2(1, N = 158) = 27.56$ , p < .001 and block two (MEAP-Y1)  $\chi^2(1, N = 182) = 78.882$ , p < .001. While the overall model itself retained significance in each block, the addition of predictors in block three (ODRs)  $\chi^2(1, N = 182) =$ 1.65, p = .198, block four (Attendance)  $\chi^2(1, N = 158) = 3.00, p = .083$ , and block five (Course Failure)  $\chi^2(1, N = 182) = .728, p < .001$  were all non-significant. Using the Nagelkerke R estimate for  $R^2$  21.4% of variance was explained in block one. Block two showed 52.4% of variance explained. The addition of variables in blocks three, four, and five showed no significant increase in variance explained. Classification results for seventh grade EWS to predict MEAP reading showed an accuracy rate of 60.1% for FRL alone in block one (see Table 4). FRL was retained in subsequent block entries. The addition of MEAP-Y1 (prior performance on MEAP reading) in block two increased classification accuracy to 74.7%. Adding ODRs in block three increased classification accuracy to 75.9%. The addition of attendance in block four produced no increase or decrease in classification accuracy. In block five the addition of course failure produced a decrease in classification accuracy down to 74.1%.

For eighth grade EWS predicting MEAP, H-L tests in all block entries were nonsignificant (see Table 3). Omnibus tests showed significance in blocks one,  $\chi^2$  (1, N = 222) = 13.16, p < .001 and two  $\chi^2$  (2, N = 221) = 123.25, p < .001. As with seventh grade, the addition of ODRs, attendance, and course failure into the model in subsequent blocks produced an overall model that was significant even though the addition of each variable was non-significant. Variance explained in block one was 7.9%. The addition of MEAP-Y1 in block two showed a significant increase in variance explained to 58.5%. Additional variables in blocks three, four, and five produced no statistically significant increase in variance explained. Classification results for 8<sup>th</sup> grade EWS to predict MEAP reading, showed a larger than expected accuracy for FRL at 64.4% in block one alone (see Table 5). The addition of MEAP-Y1 in block two increased classification accuracy to 84.7%. The addition of ODRs, attendance, and course failure in subsequent blocks produced exactly 0% increase in classification accuracy.

Logistic regression of EWS predictors for eighth grade was also done using ACT Explore<sup>®</sup> reading as the dependent variable. As a measure designed to specifically address career and college readiness ACT Explore provides external validity for the prediction model and provides a look how the prediction models function in terms of college readiness as opposed to profiency on state adopted standards. Results of EWS analysis with ACT Explore® as the dependent measure are consistent with the results of EWS analysis with MEAP reading as the dependent measure. Analyses show similarities in classification accuracy and variance explained as well as the significance of each EWS predictor in each block entry. Model fit tests for EWS predicting ACT Explore<sup>®</sup> in all block entries were non-significant (see Table 6) indicating that the models in each block were not poorly fit to the data. Omnibus tests showed significance in blocks one,  $\chi^2(1, N = 207) = 19.40$ , p = .007 and two  $\chi^2(2, N = 207) = 86.929$ , p < .001. The addition of ODRs, attendance, and course failure into the model in subsequent blocks produced an overall model that was significant though the addition of each variable was non-significant. Variance explained in block one was 12.8%. The addition of MEAP-Y1 in block two showed a significant increase in variance explained to 49.1%. Additional variables in blocks three, four, and five produced no statistically significant increase in variance explained. Classification results

for 8<sup>th</sup> grade EWS to predict ACT Explore<sup>®</sup> showed accuracy for FRL of 70.9% in block one alone (see Table 7). The addition of MEAP-Y1 in block two increased classification accuracy to 79.6%. Adding ODRs, attendance, and course failure in subsequent blocks produced an increase in classification accuracy to 81.6% though this increase was non-significant.

**Curriculum-Based Measures.** Block entry of CBM variables in the prediction model were entered in the following order: demographic variables, *easy*CBM MCRC, R-CBM, and Maze. Model fit was assessed for each block entry of variables. Test statistics for model fit and variance explained for outcomes on MEAP reading are summarized in Table 8 (seventh grade) and Table 9 (eighth grade). Chi-squared and Hosmer- Lemeshow (H-L) tests were preformed for goodness of fit.

For seventh grade CBM predicting MEAP reading, H-L tests for all five blocks were non-significant indicating that the models were not poorly fit (see table 8). Omnibus tests showed significance in block one (FRL + Ethnicity),  $\chi^2(1, N = 182) = 15.083$ , p = .002 and block two (*easy*CBM)  $\chi^2(2, N = 182) = 40.018$ , p < .001, block three (R-CBM)  $\chi^2(3, N = 182) =$ 63.374, p < .001, and block four (Maze)  $\chi^2(4, N = 182) = 64.528$ , p < .001 for the overall model. The addition of *easy*CBM in block two  $\chi^2(2, N = 182) = 25.314$  p <.001 and R-CBM  $\chi^2(3, N =$ 182) = 23.355 p < .001 in block three were significant at the point of entry and Maze in block four were significant at each point of entry  $\chi^2 = 1.15-25$ .

Nagelkerke *R* estimates of  $r^2$  showed that the increase in variance explained with the addition of each variable was significant in all four blocks. Block one showed 7.5% of variance was explained. Block two showed 38.1% of variance explained. Block three showed 51.3% variance explained. Block four showed 53.4% of variance explained. Significant increases in

variance explained and chi-squared omnibus test indicate that the each variable added significant value to the prediction model.

Results of classification accuracy for seventh grade CBM to predict MEAP reading showed a relatively high accuracy rate of 66.1% for FRL and ethnicity in block one (see Table 10). FRL and ethnicity were retained in subsequent block entries. The addition of *easy*CBM MCRC in block two increased classification accuracy to 67.8%. Adding R-CBM in block three increased classification accuracy to 73.9%. Adding Maze in block four resulted in a decrease in classification accuracy to 72.8%.

For eighth grade, CBM predicting MEAP, H-L tests in blocks one and two were nonsignificant. Block three which included *easy*CBM and R-CBM was significant H-L  $\chi^2$  (8, N = 222) = 16.1, p = .041 (see Table 9). Omnibus tests showed significance in blocks one,  $\chi^2$  (1, N = 222) = 12.525, p = .006, two  $\chi^2$  (2, N = 222) = 13.16, p < .001, three  $\chi^2$  (3, N = 222) = 104.681, p < .001, and four  $\chi^2$  (4, N = 222) = 110.119, p < .001. The addition of *easy*CBM MCRC, R-CBM, and Maze in subsequent blocks were significant at each point of entry. Variance explained in block one was 7.5%. In block two *easy*CBM MCRC showed a significant increase in variance explained to 38.1%. In block three variance explained increased to 51.3% with R-CBM added. Peak variance explained was 53.4% in block four with the addition of Maze. Classification results for 8<sup>th</sup> grade CBM to predict MEAP reading showed an accuracy of 65.9% for FRL in block one alone (see Table 11). The addition of *easy*CBM MCRC in block two increased classification accuracy to 79.6%. In block three R-CBM improved accuracy to 84.1%. For block four the addition of Maze to the model reduced overall classification accuracy to 82.7%.

When the analysis was repeated for eighth grade using ACT Explore<sup>®</sup> reading as the dependent variable, results of CBM analysis with ACT Explore<sup>®</sup> as the dependent measure were

again, consistent with the results of CBM analysis with MEAP reading as the dependent measure. Analyses show similarities in classification accuracy and variance explained as well as the significance of each EWS predictor in each block entry. Model fit tests for CBM predicting ACT Explore<sup>®</sup> in all block entries were non-significant (see Table 12) indicating that the models in each block were not poorly fit to the data. Omnibus tests showed significance for the overall model were significant (p < .005). The addition of *easy*CBM MCRC, R-CBM, and Maze variables were significant at each point of entry. Variance explained in block one for FRL and ethnicity was 9.3%. The addition of *easy*CBM MCRC in block two showed a significant increase in variance explained to 32.2%. In block four, R-CBM increased variance explained to 42.4%. In block three, Maze increased variance explained to 46.7%. Classification results for 8<sup>th</sup> grade CBM to predict ACT Explore<sup>®</sup> showed an accuracy for FRL and ethnicity of 71.1% in block one alone (see Table 13). Adding *easy*CBM MCRC in block two increased classification accuracy to 74.8%. Adding R-CBM increased accuracy to 80.7% in block three. In block four Maze increased overall accuracy to 83.5%.

#### **Linear Analysis**

To further explore the relationship between CBM, EWS, and outcomes on high stakes tests, hierarchical linear regression was conducted using MEAP reading scaled scores and ACT Explore<sup>®</sup> scaled scores as the dependent variables. Linear regression was selected in order to examine the linear relationship between each combination of predictors and dependent measures on a continuous (linear) scale. Linear regression allows for a more accurate assessment of the correlation between each prediction model and linear outcomes. Linear regression also provides the opportunity to examine a direct  $R^2$  calculation (as opposed to the Nagelkerke *R*), and the standard error of each prediction model. Scaled scores for ACT Explore and MEAP are grade

level dependent, meaning that the range of scores from not proficient to proficient in one grade do not overlap with the range of scores from another grade level. To avoid a bi-modal distribution in outcome measures, analyses were conducted separately for each grade level and each dependent variable. Analysis was done using the enter method. Block entry of each predictor variable was conducted in the same order as in the logistic regression analyses described above.

All linear regression models were examined for multicollinearity using Tolerance and Variance Inflation Factor (VIF). For the CBM models some multicollinearity was expected, given that all three CBM assessments (*easy*CBM MCRC, R-CBM, and Maze) and dependent variables (MEAP reading and ACT Explore<sup>®</sup> reading) are measures of reading performance. For EWS variables multicollinearity was also expected given that behavior problems, academic problems, and attendance problems are often comorbid. There is little evidence that multicollinearity within this investigation was at a level severe enough to compromise the interpretation of results. Tolerance and VIF were typically within acceptable limits (Tolerance > .4, VIF < 2.5) and in many cases less severe than anticipated. For seventh grade CBM assessments regressed to MEAP, reading collinearity was within acceptable limits (Tolerance > .4, VIF < 2.5) for all variables in the blocks one through three (range 1.02, 1.55). However in block four VIF exceeded 2.5 for the last two variables entered, R-CBM and Maze (range 1.06, 2.97). All other models produced test statistics for tolerance and VIF that were within the acceptable range.

**Early Warning Signs.** Using EWS to predict outcomes for scaled scores on MEAP reading for seventh grade, FRL was included in model one to control for variance due to socioeconomic status. Model one with FRL alone was statistically significant  $R^2 = .04$ , F(1,165)

= 6.95, p = .009 (see table 14). In model two the addition of MEAP-Y1 was significant  $R^2$  = .527, F(1,164) = 168.49, p < .001, indicating an increase of 48% variance explained over FRL alone. In models three through five, the addition of ODRs, attendance, and course failure were non-significant. Non-significant changes in  $R^2$  for blocks three through five range from .001 to .005 indicating that adding ODRs, attendance, and course failure to the linear combination of EWS predictors produced no significant improvement in the model.

For eighth grade, FRL was included in model one and was significant  $R^2 = .057$ , F(1,220) = 13.189, p < .001 (see Table 15). When MEAP-Y1 was added in model two, the proportion of variance explained by the model increased significantly  $R^2 = .60$ , F(2,163) = 298.688, p < .001. No other predictors added to the linear combination in subsequent models yielded a significant change to the EWS model.

Results of the linear analyses of the EWS model to predict outcomes on MEAP reading for seventh and eighth grade are consistent with results of the logistic analyses. The linear combination of demographic predictors and prior scores on MEAP reading (MEAP-Y1) accounts for the majority of variance explained in the model. ODRs, attendance, and course failure failed to significantly strengthen the model.

**Curriculum Based Measures.** Using CBM to predict outcomes for scaled scores on MEAP reading for seventh grade, model one included both ethnicity and FRL (see table 16). Model one was significant  $R^2 = .10$ , F(1,180) = 10.25, p < .001. In model two *easy*CBM MCRC was added. The combination of demographic variables and *easy*CBM MCRC produce a significant increase in variance explained  $R^2 = .32$ , F(1,179) = 55.47, p < .001. In model 3 the addition of R-CBM produced a significant increase in the percentage of variance explained by the model  $R^2 = .483$ , F(1,178) = 58.12, p < .001. The maximum variance explained was achieved

in model 4 with the addition of Maze. However, the slight increase in  $R^2$  attributed to the addition of Maze was not significant  $R^2 = .485$ , F(1,177) = .71, p = .40. These results indicate that *easy*CBM MCRC accounts for the greatest amount of variance attributable to a single predictor in the model. The addition of R-CBM improved the predictive model though comparatively less than *easy*CBM MCRC. Maze provided no added statistically significant value to the prediction model.

For eighth grade, FRL and ethnicity were included in model one which was significant  $R^2$ = .07, F(2,226) = 8.10, p < .001 (see table 17). When *easy*CBM MCRC was added in model two, the proportion of variance explained by the model increased significantly,  $R^2 = .37$ , F(1,225) =110.098, p < .001. Block three added R-CBM into the model. Results in block three produce another significant increase in variance explained  $R^2 = .495$ , F(1,224) = 53.65, p < .001. In contrast with seventh grade, the addition of Maze in model four was significant  $R^2 = .52$ , F(1,223) = 9.98, p < .001. Results for eighth grade indicated a model in which the variance is more equitably distributed amongst predictors. In the CBM model for eighth grade, all four models indicate that the new predictor introduced with each entry attributed a statistically significant amount of variance explained to the model.

As with the logistic regression analysis, linear regression analyses were repeated for eighth grade using ACT Explore<sup>®</sup> reading as the dependent variable. Linear regression showed that CBM assessments significantly predicted ACT Explore<sup>®</sup> scores similarly to previous analyses described above (see Table 18). In model one, race and ethnicity were included. Model one was significant  $R^2 = .06$ , F(2,218) = 6.64, p = .002. With *easy*CBM in model two the model was significant  $R^2 = .30$ , F(1,217) = 78.60, p < .001. In model three R-CBM was added. The results in model three were significant  $R^2 = .39$ , F(1,216) = 27.920, p < .001. In model four,

Maze was added. Again, the additional variance attributed to the addition of Maze as a predictor was significant  $R^2 = .45$ , F(1,215) = 23.829, p < .001. Results of analysis with ACT Explore<sup>®</sup> are consistent with the results of CBM analysis for outcomes on MEAP reading. It is worth noting though that in block four of CBM analysis for ACT Explore<sup>®</sup> the addition of Maze accounted for a larger amount of variance than it did in any other analysis run for this investigation.

When repeated with ACT Explore<sup>®</sup>, analysis using the EWS model produced results under linear regression that were similar to analyses using MEAP reading as the dependent variable. Model one contained FRL to control for variance related to socioeconomic status. Model one showed a significant result  $R^2 = .04$ , F(1,210) = 9.34, p = .003. In model two MEAP-Y1 was added. The resulting increase in variance explained was significant  $R^2 = .48$ , F(1,209) =173.344, p < .001. The addition of ODRs, attendance, and course failure in subsequent blocks was not significant.

Analysis using ACT Explore<sup>®</sup> confirms the results of analysis done with MEAP reading as the dependent variable. Seeing results that are consistent between state and national assessments is encouraging. The use of a national assessment as another outcome measure to crosscheck results increases confidence that results are not subject to any substantive regional effects or undesirable psychometric properties of MEAP.

When compared head-to-head, the combination of state assessment scores and EWS predictors are comparable, and in some cases more accurate than the CBM model. With MEAP reading as the dependent measure, CBM and EWS models returned comparable peak values for classification accuracy (seventh grade = CBM = 73.9%, EWS = 74.1%, eighth grade = CBM = 84.1%, EWS = 84.7%). However, EWS accounted for an overall greater amount of variance explained at seventh (55.1%) and eighth (58.1%) grade compared to CBM (seventh = 40.2%,

eighth = 53.4%). Additionally, in the EWS model fewer predictor variables were required to achieve the maximum variance explained, compared to the CBM model. For EWS, the demographic variables and prior MEAP scores accounted for the vast majority of variance explained. The addition of ODRs, attendance, and course failure added no statistically significant value to the predictive model. For CBM, the addition of *easy*CBM and R-CBM was significant at each point of entry and therefore contributed substantial value to the peak accuracy and variance explained. To put this in perspective, even with all of the CBM predictors added to the model, the CBM model did not achieve as high a value in variance explained and overall classification accuracy as did only the previous years' MEAP scores from the EWS model.

## **CHAPTER 5: Discussion**

## Summary

This investigation sought to compare two different models for the identification of students in need of reading support; one based on extant data and the other based on additional reading specific screening tests. A thorough search of extant literature revealed a need for an investigation of screening methods at the middle school level. As discussed earlier, there is currently no consensus on the most appropriate way to identify middle school students at risk of underperformance in reading. This investigation sought to bridge this gap by comparing two disparate models, one commonly used at the elementary level (CBM) and one commonly used at the high school level (EWS). Specifically, this investigation sought to answer three questions:

- How do curriculum-based measures in reading compare to early warning signs data in predicting outcomes on statewide assessments of reading at the middle school level?
- 2. What is the valued added in combining curriculum-based measures (CBM) in oral reading fluency, Maze reading comprehension, and *easy*CBM multiple choice reading comprehension as predictors of outcomes on high stakes reading assessments at the middle school level?
- 3. What is the valued added in combining Early Warning Signs (EWS) data in attendance, office discipline referrals, failing grades, and prior performance on statewide assessments function as predictors of outcomes on high stakes reading tests at the middle school level?

**Question 1.** How do curriculum-based measures in reading compare to early warning signs data in predicting outcomes on statewide assessments of reading at the middle school level?

The primary analyses of this investigation examined two important factors in model comparisons. The first is the overall classification accuracy of each model. Classification accuracy is the percentage of students that were correctly classified as being at-risk or not at-risk by each of the prediction models. The results predicted by the model are compared to students' actual scores on the outcome measures (MEAP and ACT Explore<sup>®</sup>) to yield an overall perectage of students corrected categorized by the prediction model. Comparing classification between EWS and CBM models enables a side-by-side comparison of how accurately each of these models classifies students as at-risk in reading.

The second important analysis is an examination of variance explained as defined by the Nagelgerke r (logistic regression) and  $r^2$  statistics. Variance explained is an estimate of the amount of variability in a dependent variable that can be attributed in prediction model. In the current investigation, it is helpful to examine variance explained in order to see how much of the accuracy in prediction outcomes can be attributed to each model. The greater the percentage of variance explained attributed to each model (with a p-value <.05), the more confident we can be that the model is contributing to the prediction outcomes and not a chance result.

Results were consistent across seventh and eighth grade. When MEAP was used as the outcome measure the overall classification accuracy between EWS and CBM returned a less than 1% difference (seventh grade = .2%, eighth grade = .6%). When ACT Explore<sup>®</sup> was used as the dependent variable the overall classification accuracy of EWS and CBM models were separated by 1.9% in favor of the CBM model.

The EWS model accounted for greater variance in outcomes on both MEAP and ACT Explore<sup>®</sup> than the CBM model. This is not surprising given that the EWS model included previous scores on state reading assessments. Since previous scores on statewide reading assessments are a predictor of future scores on state reading assessments (Denton, et al., 2012) it makes sense that prior state test scores account for the largest proportion of variance in the EWS model. Findings of this investigation show variance explained by past state test scores on future state test scores in reading (seventh = 55.1%, eighth =58.1%) are notably higher than results reported by Denton et al. In their analysis of scores for eighth grade students in Texas, researchers found 44% of the variance explained on future state test scores was accounted for by the previous year's state test scores. Results of Denton et al. (2012) and the current investigation support the assertion made by Reed and colleagues that data from annual statewide reading assessments should be, "an integral part of the universal screening system," for reading in secondary schools (Reed, Wexler, & Vaughn, 2012, p. 49).

Given that state reading assessments are a measure of reading, one might posit that state assessment scores are more aligned with the constructs assessed in the CBM model than those measured in the EWS model. Thus, state reading tests and CBM should be grouped together as a collective set of reading data. However, state reading tests do, in fact, assess a dimension of reading that is notably different than CBM universal screeners. State assessments of reading are a measure of overall grade level reading achievement aligned with state adopted standards. CBM measure more narrow skills such as oral reading fluency, which is then used as a proxy for reading achievement. Putting state reading assessments and CBM reading assessments into the prediction model may provide a more powerful overall prediction model with greater accuracy and reliability. Further analysis exploring the use of state assessment data and CBM reading data

to predict which students are in need of reading intervention is recommended. In addition, examination of state assessment data and each CBM measure independently provides information that teachers can use to pinpoint specific areas that warrant reading support. For example, comparing scores on R-CBM (oral reading fluency) and *easy*CBM MCRC (reading comprehension) may help teachers prioritize intervention services specific to students' individual needs.

However, it is important to keep in mind that the grouping of variables in each prediction model was based on the distinction between data that schools are already required to collect based on federal and state requirements (i.e. statewide achievement tests, attendance records, discipline records, and course grades) and assessments that must be administered in addition to state mandated testing requirements, formative, and summative content area tests. The impetus for this investigation comes from the underlying question, Can schools use the data they already have to identify which students need reading intervention services or must they do additional testing? In the case of the school participating in this study, results support using state test data as the best screening measure and the one that should be used at a true universal level (i.e., for all students), while reserving CBM screening when statewide test data are not available. The one exception to using state assessment data as a truly universal screener is that students exempt from MEAP testing based on their IEP qualifications will not have state reading test scores.

**Question 2.** What is the valued added in combining curriculum-based measures (CBM) in oral reading fluency, Maze reading comprehension, and *easy*CBM multiple choice reading comprehension as predictors of outcomes on high stakes reading assessments at the middle school level?

Results of both the binary and linear CBM analysis indicate that *easy*CBM accounted for the largest proportion of variance attributable to the CBM measures. This is perhaps not surprising given that *easy*CBM, and the dependent measures used in this investigation (MEAP reading and ACT Explore<sup>®</sup>) are both multiple choice reading tests that assess factual, inferential, and evaluative questions. While the analysis in this investigation did not explore the effects of test format as a moderator of the relationship between dependent and independent variables there may be some advantages to using a multiple choice reading test to predict outcomes on another multiple choice reading test. Presumably, multiple choice reading comprehension assessments all assess similar skills and constructs and therefore would be expected to correlate with one another more strongly than those that assess different aspects of reading such as fluency or decoding.

The addition of R-CBM provided an incremental increase that was small but consistently significant across grade levels, dependent measures, and analyses. As a measure of oral reading fluency R-CBM measures a different aspect of reading than does *easy*CBM. While *easy*CBM focuses on passage reading, recall, and application, R-CBM is a test of speed and accuracy. The fact that both assessments contributed significantly to the variance attributed to the CBM predictive model means that neither one is likely a comprehensive measure of reading ability. Instead, *easy*CBM and R-CBM measure specific aspects of reading that serve as a proxy for reading achievement as a whole.

Including Maze in the model failed to produce consistent or statistically significant contributions to the CBM model. Though the lack of consistent positive results with Maze may initially seem problematic in the model, it is important to consider that Maze is a combined measure of silent reading fluency and reading comprehension. Since the two variables entered immediately prior to Maze in the model are measures of reading comprehension (*easy*CBM) and

fluency (R-CBM) it is entirely possible that variance potentially attributable to Maze was already accounted for by R-CBM and *easy*CBM. It is also possible that the lack of significant effect on the prediction model with the addition of Maze was a consequence of the strength of the relationship between Maze and the dependent measures. As noted in Chapter Two, prior research has shown that the relation between established measures of reading comprehension and Maze is not as strong it is between these same external measure and ORF (Ardoin et al., 2004; Parker, Hasbrouk, & Tindal, 1992; Tolar, et al., 2012).

Additionally, data from the CBM model at seventh grade indicated a discrepancy between the consistency of classification accuracy and variance explained. For seventh grade CBM analysis using MEAP as the dependent variable, the variance explained by the full model is 40.2%, which is well below all other percentages of variance explained in this investigation across all models and grades (range 46.7%-58.7%). This is somewhat surprising given that the classification accuracy and between the EWS (74.1%) and CBM (72.8%) are within 2% while there is nearly a 15% discrepancy in variance explained between for EWS (55.1%) and CBM (40.2%). Staff from the participating school indicated that there are often anomalies in seventh grade CBM data trends that are attributed to the comparatively high cut scores established at seventh grade. Further investigation of seventh grade CBM data is necessary to more fully understand the differences in variance explained between the CBM and EWS models.

**Question 3.** What is the valued added in combining Early Warning Signs (EWS) data in attendance, office discipline referrals, failing grades, and prior performance on statewide assessments function as predictors of outcomes on high stakes reading tests at the middle school level?

For the EWS models, it is clear that prior scores on state assessments in reading account for the vast majority of variance explained on future outcomes of state reading assessments. This pattern occurred at both seventh and eighth grade for outcomes on MEAP and ACT Explore<sup>®</sup>. These findings are consistent with similar analysis conducted by Denton and colleagues (2011) that found prior scores on state reading assessments in Texas to be the best overall predictor of future scores on state reading assessments compared to nine other commonly used reading assessments.

The addition of ODRs, attendance, and course failure into the models produced only incremental non-significant increases in variance explained and classification accuracy. Typically, these results would indicate that ODRs, attendance, and course failure provide little to no added value for the prediction model. These results are somewhat surprising given the reported strength of behavior, attendance, and course failure as predictors of high school dropout (Balfanz, 2009; Balfanz, Herzog, & Mac Iver, 2007; Neild, Balfanz, & Herzog, 2007). The key to the weak value added by attendance, behavior, and course failure may be in part due to comorbidity of academic and behavior problems. That is, because variables included in early warning signs appear to be functionally related, students that struggle in one area are also likely to struggle in another. For example, students with poor attendance (>10% absences) are more likely to fail courses than those with adequate attendance. This means there is likely a considerable amount of overlap in the groups of students predicted at-risk if the EWS variables were each explored independently. Just as in the CBM results, the amount of variance predicted in the EWS model by prior state test scores could include some or all of the potential variance explained by ODRs, attendance, and course failure. Results of the investigation reported above

by Neild, Balfanz, & Herzog (2007) elude to this in reporting that the chances of a student dropping out of high school was 75% for students that exhibiting one or more warning signs.

It is important to consider though the balance between the value added by including ODRs, attendance, and course failure in the model and the costs associated with collecting and managing such data. Since these three data come from extant sources, require no additional testing, are routinely collected, and are often required for state reporting purposes, it may be beneficial to keep ODRs, attendance, and course failure as a part of screening procedures. Therefore the incremental benefits in accuracy of prediction may outweigh the time, effort, and resources necessary to include these data. Furthermore, even non-significant increases in prediction accuracy leads to better overall efficiency in identification of students at-risk and reduces the potential for false positive and false negative predictions. For the participating school, consider that a 2% increase in overall accuracy translates into approximately seven more students correctly identified as needing or not needing reading intervention services. In sum, if schools use an EWS model for identification of students in need of reading support, and the data in ODRs, attendance, and course failure are readily available, school personnel should consider such data for screening purposes.

## Limitations

Data for this investigation came from a single school serving two grade levels (seventh and eighth). The sample size (n = 434) for both grades combined is not atypical in the field of CBM research (Yeo, 2009), but nonetheless could benefit from a larger sample size. Because data were limited to a single site it was not possible to control for school-level, district level, or regional level effects in the analysis. However, as evidence in table 20, there are several areas in which the participating school is representative of the larger population. In school-wide

enrollment, the participating school includes 14% students with disabilities as a percentage of total enrollment, which is comparable to both state (13%) and national (13%) public school enrollment. Likewise, the percentage of students qualifying for free or reduced lunch at the participating school (53%) is similar to state (47%) and federal (48%) levels. On important area in which the school, state, and federal profiles are quite different is in the area of reading achievement. The percentage of students proficient in reading as measured by MEAP is 65% for student in 8<sup>th</sup> grade. This is 8% below the statewide average (73%). When compared nationally, results of National Assessment of Educational Progress (NAEP) show 36% of eighth grade students proficient in reading. Since the primary are of interest for this investigation is reading, it is important to recognize that results are generalizable to schools with similar demographic and achievement characteristics rather than the population as a whole. Future research should include data with a larger sample size across multiple schools, districts, and regions.

Another important limitation in the current study is the use of MEAP and ACT Explore<sup>®</sup> as dependent variables. As mentioned above, MEAP and ACT Explore<sup>®</sup> have both been discontinued by their respective organizations beginning with the 2014-2015 school year. With the development of Common Core State Standards (CCSS) and computer adaptive testing, ACT, the Michigan Department of Education, and many other state departments of education have transitioned to tests that are more closely aligned with CCSS and incorporate technology enhanced test times such as drag and drop items, hot spots (e.g., clicking on a target area or highlighted area of text), short answer, and extended response items available in online testing environments. MEAP has been replaced by M-STEP as the statewide test of academic achievement and school accountability in Michigan. ACT Explore<sup>®</sup> has been replaced by ACT

Aspire<sup>®</sup>. While the results of the current investigation are informative, it will not be possible to replicate these results since MEAP and ACT Explore<sup>®</sup> are no longer available.

Another limitation is the exclusive use of secondary data. Data were provided to the researcher directly from the participating school and it is not possible to assess the fidelity of testing procedures and the accuracy of data entry. A study on training and fidelity of test procedures conducted by Reed and Sturges found that 8% of all ORF administrations resulted in procedural errors that compromised the data (Reed & Sturges, 2012). Similary, Cummings, Biancarosa, Scchaper and Reed (2014) found that up to 16% of variance in ORF scores can be attributed to variation between test proctors. While there is no evidence in the current study of poor implementation fidelity that would compromise the findings, ensuring fidelity of investigation.

Finally, it is important to recognize that this analysis of EWS and CBM focuses only one of the many functions of these data sources, and that is universal screening. CBM and EWS data can be useful for other aspects of assessment including progress monitoring, evaluation of instructional effectiveness, and as part of the referral process for special education, to name a few. This investigation explored the use of EWS and CBM as methods for distinguishing between those students that do and do not require additional reading instruction and intervention. Though EWS and CBM appear to be comparable methods for identification purposes, there was no investigation of how these models inform what happens after students have been identified as needing additional services. Further investigation is needed to know if data used in the screening process actually enables educators to appropriately fit students with the targeted services they need. Between EWS and CBM there may be advantages in using one over the other when it

comes to deciding what services are the best fit for a student including the specific instructional program, frequency, duration, and group size. Beyond identification, it is important for schools to know how best to proceed with students that could benefit from further intervention.

## Implications

The current investigation represents a departure for prior students in EWS and CBM in two critical ways. First, the latency between the CBM data collected and the primary dependent measure is markedly short. Since MEAP testing was completed in the fall of each academic year, the fall benchmark cycle of CBM testing concluded less than three weeks before the start of MEAP administration. The reduced latency between testing periods effectively eliminates the intervention and maturation effects found in students that compare fall CBM data to state test data from a spring administration. Additionally, the EWS data used in these analyses were collected from the prior year. One might imagine that the close proximity between CBM and MEAP compared the EWS data and MEAP would give the CBM model a distinct advantage in its ability to accurately classify students. While reducing latency between CBM may have been advantages in predicting outcomes, the magnitude of such advantage was not such that it surpassed the overall classification accuracy of the EWS model.

Second, much of the work done in curriculum-based measurement as well as other areas of educational measurement assumes that assessment is required to gather the required data for decision making, whether it be at the student, teacher, grade level, school, or district levels. However, it may be the case that the data required to make sound educational decisions may already exist. If students can be accurately categorized as at-risk of failure in reading without the use of additional tests, there is potential to create systems of assessment, decision-making, and support that are more efficient and effective.

Though the results of the current investigation show that EWS and CBM accurately categorize students as at-risk or not at-risk for failure on high stakes tests at similar rates, it should be noted that results do not support the elimination of CBM testing in middle schools in favor of EWS. It may be tempting to compare the classification accuracy of EWS versus CBM and conclude that CBM are no longer required for screening purposes. This may be particularly tempting to conclude when educators already question the number of tests administered to students each year. However, there are likely to be situations in which the use of CBM are necessary for the identification of students in need of support. These situations include data that are inaccurate or incomplete due to frequent absences, data that are inaccessible because of changes in building enrollment, or for those students whose archival data are inaccessible or simply do not exist. In these cases, the use of CBM as a screen method is recommended.

There are also many other functions of CBM that were not addressed in this investigation. CBM can be used for weekly and monthly monitoring of students' progress over time, evaluate the performance of an individual or class relative to local, state, and national norms of performance, and evaluate an individual student's rate of growth and response to instruction (Salvia, Ysseldyke, & Bolt, 2012). Even if extant data are used for screening purposes, CBM should remain a part of a well-balanced school-wide system of assessment.

It is also important to mention that there are several barriers that may inhibit schools' ability to use extant data as a method for universal screening. Chief among these barriers is access to extant data. In many cases teachers, interventionists, and support staff may not have direct access to students' past records of state assessment scores, absences, ODRs, and course failure. In accessibility can occur for a variety of reasons including delays in transfers of records from previous schools and inadequate credentials for accessing such data on district student

information systems. Likewise, schools may not have sufficient data infrastructure and/or expertise necessary to compile EWS data in a format that is suitable for screening purposes. While the calculations required to aggregate ODRs, attendance, and course failure require only rudimentary arithmetic, doing so *en masse* class-wide or building-wide can be time consuming and logistically challenging. In some cases, it may be more efficient and cost effective to administer CBM than it is to manage and maintain EWS data. There are certainly technological options that make compiling EWS data less challenging, however implementation of technological solution requires additional time, training, infrastructure, and funding.

Finally, the exploration of extant and CBM data to predict outcomes on high stakes tests for students in middle schools need not be limited to an either/or scenario. Though analyses suggests that EWS and CBM assess risk status in reading for students in seventh and eighth grade with a similar level of accuracy, the most accurate and efficient method of identifying students in need of support may in fact come from a combination of EWS and CBM assessment. As suggested above, a blend of state assessment data and CBM screening is one promising option. Another potential option for screening is a tiered system of screening in which EWS data (including state assessment data) are used as the primary screening instrument. CBM assessments could then be used as a secondary level screening tool for students whose EWS data is at or near predetermined benchmarks. Furthermore, including measures of oral reading fluency and reading comprehension to EWS may assist school personnel in differentiating between students that need targeted support in a specific area or comprehensive reading support.

### **Suggestions for Future Research**

Results of this investigation suggest the need for further research in several areas. Chief among these is the need for replication using results from Common Core State Standards (CCSS)

assessments including Smarter Balanced Assessments Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC). As states transition to CCSS aligned state assessments it is necessary to understand if the EWS and CBM will function equally well in predicting performance new high stakes tests, in particular those using online administration and technology enhanced assessment items.

Another area requiring further research is the exploration of potential bias in EWS and CBM data for disaggregated subgroups in socioeconomic status, race/ethnicity, and disability classification. Prior research suggests that students in many of these subgroups are at higher risk of underperformance and high-school dropout than the aggregated population (Hosp, Hosp, & Dole, 2011). It is therefore important to ensure that screening measures are equally accurate, or better, at identifying students in need of support for at-risk populations as they are for the general population. Though some investigation of potential bias among CBM measures by subgroup has been done, there are still many answered questions particularly for students in middle school (DeBoy, 2013; Hosp, Hosp, & Dole, 2011).

Results of this study suggest that extant data sources, including prior scores on statewide reading achievement tests, are a feasible alternative to the use of curriculum-based measures in reading for universal screening of students in middle school. However it is presently unclear how the prediction models tested in this study will play out in practice for school personnel. Arming schools with accurate and efficient methods for identification of students in need of support is only the first step. The ways in which school personnel use CBM and EWS data to make decisions regarding services for students and allocation of resources is equally important. Future research is needed that explores the practical implications and feasibility of using CBM and EWS as a mechanism for universal screening in middle schools.

Finally, it is essential to recognize much of the current research in screening methods at the middle school level assumes that a valid and reliable method for risk identification already exists amongst the many assessments and systems available. However, it is entirely possible that the most accurate and reliable screening methods in middle school may not yet exist. The uniqueness of adolescents in their academic, behavioral, and social-emotional growth may warrant a set of screening tools that are demonstrably different than anything that currently exists. Further research is needed to explore creative methods for screening that meets the specific needs of adolescent learners. APPENDIX

							Shapiro-Wilk		
Measure	Grade	$\bar{x}$	Md	SD	Skewness	Kurtosis	Test	df	р
easyCBM MCRC	7	11.82	12	3.73	483	556	.951	192	<.001
	8	13.36	14	3.75	-1.00	1.032	.928	235	<.001
R-CBM	7	137.06	136	37.84	190	.125	.990	194	.229
	8	145.91	149	38.40	.108	.998	.987	233	.005
Maze	7	24.41	25	8.85	.116	066	.992	192	.429
	8	25.57	26	8.79	.040	.099	.992	232	.264
ACT Explore <sup>®</sup> Scaled Score	8	13.7	13	3.734	.003	417	.956	228	.001
MEAP	7	721.61	720	33.15	.712	1.423	.955	196	<.001
Reading Scaled Score	8	822.23	824	24.20	001	.314	.986	242	.020

Table 1Summary of Descriptive Statistics and Tests for Normality

*Note.* Md = median. SD = standard deviation. MCRC=Multiple Choice Reading Comprehension. R-CBM=Reading Curriculum Based Measure.
		Omn	ibus T	est	H-	L Test			
		$\chi^2$	df	р	$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
	Step	27.560	1	<.001					
Gender+FRL+Ethnicity	Block	27.560	1	<.001	2.452	8	.964	191.246	.214
	Model	27.560	1	<.001					
	Step	71.905	1	<.001					
MEAP-Y1*	Block	71.905	1	<.001	6.145	8	.631	139.924	.524
	Model	78.882	2	<.001					
	Step	1.654	1	.198					
MEAP-Y1 + ODR*	Block	1.654	1	.198	6.964	8	.533	138.270	.533
	Model	80.536	3	<.001					
	Step	3.003	1	.083					
MEAP-Y1 + ODR +	Block	3.003	1	.083	13.506	8	.096	135.267	.548
Attendance*	Model	135.267	4	<.001					
	Step	.729	1	.393					
MEAP-Y1 + ODR +	Block	.729	1	.393	12.433	8	.133	134.538	.551
Attendance + Course Failure*	Model	134.538	5	<.001					

Table 2Model Fit for Seventh Grade EWS Variables on MEAP Reading.

*Note.* Classification cutoff = 0.5 \*Gender and Ethnicity were non-significant (p > .05) in block 1, FRL was retained in all subsequent block entry of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MEAP-Y1 = Michigan Education Assessment Program reading subtest. ODR = Office Discipline Referral. Attendance = Total number of period absences. Course Failure = Total number of courses failed. *n* = 158.

		Omn	ibus T	est	H-	L Test			
	-	$\chi^2$	df	р	$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
	Step	13.160	1	<.001					
Gender+FRL+Ethnicity	Block	13.160	1	<.001	.000	0	NA	275.883	.079
	Model	13.160	1	<.001					
	Step	110.092	1	<.001					
MEAP-Y1*	Block	110.092	1	<.001	7.519	8	.452	165.791	.585
	Model	123.251	2	<.001					
	Step	.160	1	.689					
MEAP-Y1 + ODR*	Block	.160	1	.689	9.851	8	.276	165.631	.586
	Model	123.411	3	<.001					
	Step	.278	1	.598					
MEAP-Y1 + ODR +	Block	.278	1	.598	8.515	8	.385	165.354	.587
Attendance*	Model	123.689	4	<.001					
	Step	.060	1	.806					
MEAP-Y1 + ODR +	Block	.060	1	.806	11.034	8	.200	165.293	.587
Attendance + Course Failure*	Model	123.749	5	<.001		-			

Table 3Model Fit for Eighth Grade EWS Variables on MEAP Reading.

*Note.* Classification cutoff = 0.5. \*Gender and ethnicity were non-significant (p > .05) in block 1, FRL was retained in all subsequent block entry of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. ODR = Office Discipline Referral. Attendance = Total number of period absences. Course Failure = Total number of courses failed. n = 229.

		Prof	icient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{eta}$	c for e	<sup>8</sup> (95%)
FRL	No	51	25	67.1	.860	.330	6.800	1	.009	2.362	1.238	4.507
	Yes	38	44	53.7								
	Overal % Correct			60.1								
MEAP-Y1	No	57	19	75.0	.087	.015	35.567	1	<.001	1.060	1.060	1.122
	Yes	21	61	74.4								
	Overal % Correct			74.7								
MEAP-Y1	No	56	20	73.7	.087	.015	34.843	1	<.001	1.091	1.060	1.122
ODR	Yes	18	64	78.0	109	.087	1.542	1	.214	.897	.756	1.065
	Overal % Correct			75.9								
MEAP-Y1	No	54	22	71.1	.093	.016	34.467	1	<.001	1.098	1.064	1.133
ODR	Yes	16	66	80.5	116	.089	1.718	1	.190	.890	.748	1.059
Attendance	Overal % Correct			75.9	.023	.013	3.071	1	.080	1.023	.997	1.049
MEAP-Y1	No	53	23	69.7	.090	.016	30.427	1	<.001	1.094	1.060	1.129
ODR	Yes	18	64	78.0	094	.093	1.036	1	.309	.910	.759	1.091
Attendance	Overal % Correct			74.1	.023	.013	3.175	1	.075	1.023	.998	1.049
Course Failure					152	.184	.682	1	.409	.859	.599	1.232

Table 4Logistic Regression Results for Seventh Grade EWS on MEAP Reading.

*Note.* Classification Cutoff = 0.5, MEAP-Y1 = Michigan Education Assessment Program reading subtest, ODR = Total number of Office Discipline Referrals, Attendance = Total number of period absences, FRL = Free and Reduce Price Lunch. n = 158.

		Profi	cient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{\beta}$	c for $e^{\beta}$	(95%)
FRL	No	0	79	.0	1.038	.292	12.62	1	<.001	2.822	1.592	5.003
	Yes	7	143	100								
	Overal % Correct			64.4								
MEAP-Y1	No	60	19	75.9	.080	.011	51.922	1	<.001	1.083	1.060	1.107
	Yes	15	128	89.5								
	Overal % Correct			84.7								
MEAP-Y1	No	59	20	74.7	.080	.011	51.882	1	<.001	1.084	1.060	1.108
ODR	Yes	15	128	89.5	.016	.041	.156	1	.693	1.016	.938	1.100
	Overal % Correct			84.2								
MEAP-Y1	No	59	20	74.7	.080	.011	51.333	1	<.001	1.084	1.060	1.108
ODR	Yes	15	128	89.5	.019	.041	.223	1	.636	1.020	.941	1.105
Attendance	Overal % Correct			84.2	001	.002	.274	1	.601	.999	.996	1.003
MEP-Y1	No	60	19	75.9	.080	.011	49.489	1	<.001	1.083	1.059	1.108
ODR	Yes	15	128	89.5	.024	.045	.283	1	.595	1.024	.937	1.120
Attendance	Overal % Correct			84.7	001	.002	.122	1	.726	.999	.995	1.003
Course Failure					022	.090	.060	1	.806	.978	.820	1.167

Table 5Logistic Regression Results for Eighth Grade EWS on MEAP Reading.

*Note.* Classification Cutoff = 0.5, MEAP-Y1 = Michigan Education Assessment Program reading subtest, ODR = Total number of Office Discipline Referrals, Attendance = Total number of period absences, FRL = Free and Reduce Price Lunch. n = 229.

		Omr	ibus T	est		Н	-L Test			
	-	$\chi^2$	df	р		$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
	Step	19.396	1	.007						
Gender+FRL+Ethnicity	Block	19.396	1	.007		.000	0	NA	229.154	.128
	Model	19.396	1	.007						
	Step	78.169	1	<.001						
MEAP-Y1*	Block	78.169	1	<.001	-	7.277	8	.507	161.622	.491
	Model	86.929	2	<.001						
	Sten	1 633	1	201						
MEAP-Y1 + ODR*	Block	1.633	1	201	C	9 462	8	305	159 989	499
	Model	88.561	3	<.001		. 102	0	.505	109.909	.199
	Step	2 241	1	138						
MEAP-Y1 + ODR +	Block	2.241	1	138		3 1 1 6	8	927	157 748	509
Attendance*	Model	90.802	4	<.001	-		Ũ	.,_,	1071710	.007
	Step	1.179	1	.278						
MEAP-Y1 + ODR +	Block	1.179	1	.278	2	3.321	8	.913	159.569	.514
Attendance + Course Failure*	Model	91.982	5	<.001	-		-			

Table 6Model Fit for Eighth Grade EWS Variables on ACT Explore<sup>®</sup> Reading.

*Note.* Classification cutoff = 0.5 \*Gender and Ethnicity were non-significant (p > .05) in block 1, FRL was retained in all subsequent block entry of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. ODR = Office Discipline Referral. Attendance = Total number of period absences. Course Failure = Total number of courses failed. *n* = 206.

		Profi	cient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{eta}$	c for e <sup>4</sup>	<sup>8</sup> (95%)
FRL	No	146	0	100	.934	.323	8.337	1	.004	2.544	1.350	4.796
	Yes	60	0	0.0								
	Overal % Correct			70.9								
MEAP-Y1	No	130	16	89.0	.071	.011	40.841	1	<.001	1.074	1.051	1.098
	Yes	26	34	56.7								
	Overal % Correct			79.6								
MEAP-Y1	No	131	15	89.7	.071	.011	40.253	1	<.001	1.073	1.050	1.097
ODR	Yes	25	35	58.3	076	.062	1.489	1	.222	.927	.821	1.047
	Overal % Correct			80.6								
MEAP-Y1	No	131	15	89.7	.074	.012	39.947	1	<.001	1.077	1.052	1.102
ODR	Yes	24	36	60.0	095	.062	2.321	1	.128	.910	.806	1.027
Attendance	Overal % Correct			81.1	.005	.003	2.333	1	.127	1.005	.999	1.011
MEP-Y1	No	132	14	90.4	073	012	37 815	1	< 001	1 075	1 051	1 101
ODR	Yes	24	36	60.0	077	.064	1.470	1	.225	.926	.817	1.049
Attendance	Overal % Correct			81.6	.006	.003	3.158	1	.076	1.006	.999	1.013
Course Failure					168	.166	1.020	1	.312	.845	.610	1.171

Table 7Logistic Regression Results for Eighth Grade EWS on ACT Explore<sup>®</sup> Reading.

*Note.* Classification Cutoff = 0.5, MEAP-Y1 = Michigan Education Assessment Program Reading Subtest, ODR = Total number of Office Discipline Referrals, Attendance = Total number of period absences, FRL = Free and Reduce Price Lunch. n = 206.

		Omr	ibus T	est	H	-L Test			
	-	$\chi^2$	df	р	$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
	Step	15.083	1	.002					
Gender+FRL+Ethnicity	Block	15.083	1	.002	11.878	8	.157	234.361	.107
	Model	15.083	1	.002					
	Step	25.314	1	<.001					
easyCBM MCRC	Block	25.314	1	<.001	11.312	8	.185	209.426	.266
	Model	40.018	2	<.001					
	Step	23.355	1	<.001					
easyCBM MCRC + R-CBM	Block	23.355	1	<.001	4.149	8	.843	186.070	.396
	Model	63.374	3	<.001					
	Step	1.154	1	.283					
easyCBM MCRC + R-CBM +	Block	1.154	1	.283	5.283	8	.727	184.916	.402
Maze	Model	64.528	4	<.001					

### Table 8Model Fit for Seventh Grade CBM Variables on MEAP Reading.

*Note.* Classification cutoff = 0.5 \*Gender was non-significant (p > .05) in block 1, Ethnicity and FRL wer retained in all subsequent block entries of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC=Multiple Choice Reading Comprhension. R-CBM=Reading Curriculum Based Measure. n = 180.

		Omn	ibus T	est	H	-L Test			
		$\chi^2$	df	р	$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
	Step	12.525	1	.006					
Gender+FRL+Ethnicity	Block	12.525	1	.006	1.766	8	.987	277.434	.075
	Model	12.525	1	.006					
	Step	61.031	1	<.001					
easyCBM MCRC	Block	61.031	1	<.001	3.854	8	.870	217.111	.381
	Model	72.849	2	<.001					
	Step	31.832	1	<.001					
<i>easy</i> CBM MCRC + R-CBM	Block	31.832	1	<.001	16.100	8	.041	185.279	.513
	Model	104.681	3	<.001					
	Step	5.438	1	.020					
<i>easy</i> CBM MCRC + R-CBM +	Block	5.438	1	.020	7.213	8	.514	179.841	.534
Maze	Model	110.119	4	<.001					

### Table 9Model Fit for Eighth Grade CBM Variables on MEAP Reading.

*Note.* Classification cutoff = 0.5 \*Gender and ethnicity were non-significant (p > .05) in block 1, FRL was retained in all subsequent block entries of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC=Multiple Choice Reading Comprhension. R-CBM=Reading Curriculum Based Measure. n = 226.

		Profi	cient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{eta}$	c for $e^{b}$	<sup>8</sup> (95%)
FRL	No	65	27	70.7	.676	.315	4.592	1	.032	1.965	1.059	3.646
Ethnicity	Yes	34	54	61.4								
	Overal % Correct			66.1								
easyCBM	No	65	27	70.7	.087	.015	35.567	1	<.001	1.060	1.060	1.122
MCRC	Yes	31	57	64.8								
	Overal % Correct			67.8								
easyCBM	No	69	23	75.0	.097	.058	2.830	1	.093	1.102	.984	1.235
MCRC	Yes	24	64	72.7	.032	.007	19.036	1	<.001	1.033	1.018	1.048
R-CBM	Overal % Correct			73.9								
<i>easy</i> CBM MCRC	No	69	23	75.0	.087	.059	2.212	1	.056	1.091	.973	1.225
R-CBM	Yes	26	62	70.5	.026	.009	8.434	1	.004	1.027	1.009	1.045
Maze	Overal % Correct			72.8	.037	.034	1.140	1	.286	1.037	.970	1.110

Table 10Logistic Regression Results for Seventh Grade CBM on MEAP Reading.

*Note.* Classification cutoff = 0.5 \*Gender was non-significant (p > .05) in block 1, Ethnicity and FRL were retained in all subsequent block entries of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. *n* = 180.

		Prof	icient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{eta}$	c for e <sup>t</sup>	<sup>3</sup> (95%)
FRL	No	0	77	0	989	294	11 300	1	001	2.688	1 510	4 784
1112	Yes	0	149	100	., 0,	, .	11.500	1	.001	2.000	1.010	1.701
	Overal % Correct	-	- 12	65.9								
easyCBM	No	47	30	61.0	.377	.059	41.384	1	<.001	1.458	1.300	1.635
MCRC	Yes Overal % Correct	16	133	89.3 79.6								
easvCBM	No	54	23	70.1	.290	.062	22.158	1	<.001	1.337	1.185	1.509
MCRC R-CBM	Yes Overal % Correct	13	136	91.3 84.1	.033	.007	24.232	1	<.001	1.033	1.020	1.047
		50	24	(0,0)	2(0	0(0	10.460	1	< 0.0.1	1 207	1 1 5 7	1 477
easyCBM MCRC	No	53	24	68.8	.268	.062	18.462	1	<.001	1.307	1.15/	1.4//
R-CBM	Yes	15	134	89.9	.023	.008	9.061	1	.003	1.023	1.008	1.039
Maze	Overal % Correct			82.7	.075	.033	5.243	1	.022	1.078	1.011	1.150

Table 11Logistic Regression Results for Eighth Grade CBM on MEAP Reading.

*Note.* Classification cutoff = 0.5 \*Gender and ethnicity were non-significant (p > .05) in block 1, FRL was retained in all subsequent block entries of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. n = 226.

		Omn	ibus T	est	Η	-L Test	Ţ		
	-	$\chi^2$	df	р	$\chi^2$	df	р	-2 Log Likelihood	Nagelkerke R
Gender+FRL+Ethnicity	Step Block Model	14.740 14.740 14.740	1 1 1	.002 .002 .002	6.004	8	.647	247.403	.093
easyCBM MCRC*	Step Block Model	42.051 42.051 55.658	1 1 3	<.001 <.001 <.001	6.670	8	.573	206.485	.322
easyCBM MCRC + R-CBM*	Step Block Model	21.151 21.151 76.809	1 1 4	<.001 <.001 <.001	10.549	8	.229	185.334	.424
<i>easy</i> CBM MCRC + R-CBM + Maze*	Step Block Model	9.532 9.532 86.341	1 1 5	.002 .002 <.001	4.353	8	.824	175.802	.467

Table 12Model Fit for Eighth Grade CBM Variables on ACT Explore® Reading.

*Note.* Classification cutoff = 0.5 \*Gender was non-significant (p > .05) in block 1, FRL and Ethnicity were retained in all subsequent block entry of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. n = 207.

		Profi	cient									
Predictors	Observed	No	Yes	% Correct	β	$SE \beta$	Wald's $\chi^2$	df	р	$e^{eta}$	c for e <sup>4</sup>	<sup>3</sup> (95%)
FRL	No	155	0	100	.862	.316	6.850	1	.009	2.284	1.231	4.241
Ethnicity	Yes	63	0	0	171	.083	4.190	1	.041	.843	.719	.993
	Overal % Correct			71.1								
EasyCBM	No	139	16	89.7	.409	.078	27.614	1	<.001	1.506	1.292	1.754
MCRC	Yes	39	24	38.1								
	Overal % Correct			74.8								
EasyCBM	No	143	12	92.3	.342	.084	16.633	1	<.001	.850	.704	1.026
MCRC	Yes	30	33	52.4	.027	.007	17.036	1	<.001	1.027	1.014	1.041
R-CBM	Overal % Correct			80.7								
EasyCBM MCRC	No	142	13	91.6	.311	.086	13.229	1	<.001	1.365	1.154	1.615
R-CBM	Yes	23	40	63.5	.018	.007	6.174	1	.013	1.018	1.004	1.032
Maze	Overal % Correct			83.5	.093	.031	8.725	1	.003	1.097	1.032	1.166

Table 13Logistic Regression Results for Eighth Grade CBM on ACT Explore<sup>®</sup> Reading.

*Note.* Classification cutoff = 0.5 \*Gender was non-significant (p > .05) in block 1, FRL and Ethnicity were retained in all subsequent block entry of variables. FRL = Students qualifying for free and reduced price school lunch. H-L=Hosmer-Lemeshow. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. n = 218.

	Model 1				Model 2			Model 3			Model 4			Model 5		
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	
FRL	-13.493	5.118	201	-3.495	3.687	052	-3.685	5 3.681	055	-4.118	3.745	061	-3.878	3.753	058	
MEAP-Y1				.862	.066	.713**	.858	.066	.709**	.859	.066	.710**	.841	.069	.695**	
ODR							-1.117	.835	072	-1.140	.837	073	883	.876	057	
Total Absences Course Failure										.042	.064	.036	.049 -1.313	.064 1.322	.042 059	
R <sup>2</sup>		.040**			.527**			.532			.533			.536		
F for change in R <sup>2</sup>		6.950**	:		168.489**			1.789			.437			.986		

Table 14Linear Regression Results for Seventh Grade EWS on MEAP Reading.

*Note:* MEAP-Y1 = Michigan Education Assessment Program reading subtest for previous year. FRL = Students qualifying for free or reduced price lunch. ODR = Total Office Discipline Referrals. Total Absences = Total number of period absences. Course Failure = Total number of courses failed \*p < .05. \*\*p < .01. n = 165.

#### Table 15

Model 1					Model 2			Model 3			Model 4	1	Model 5		
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β
FRL	-15.637	4.306	- .238* *	-3.329	2.896	051	-3.285	2.896	050	-3.339	2.922	051	-3.440	2.930	052
MEAP-Y1				.846	.049	.761* *	.837	.050	.753**	.838	.050	.754**	.848	.052	.763**
ODR							333	.320	045	343	.327	047	463	.372	063
Total Absences										.002	.015	.007	001	.016	004
Course Failure													.488	.720	.039
$R^2$		.057**			.601**			.603			.603			.604	
F for change in $\mathbb{R}^2$	-	13.189**	*	2	298.688**	*		1.084			.025			.459	

*Linear Regression Results for Eighth Grade EWS on MEAP Reading.* 

*Note:* MEAP-Y1 = Michigan Education Assessment Program reading subtest for previous year. FRL = Students qualifying for free or reduced price lunch. ODR = Total Office Discipline Referrals. Total Absences = Total number of period absences. Course Failure = Total number of courses failed \*p < .05. \*\*p < .01. n = 221.

		Model 1			Model 2			Model 3		Model 4			
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	
FRL	-11.367	4.778	169**	-7.960	4.211	119	-4.023	3.703	060	-4.268	3.717	064	
Ethnicity	-4.052	1.151	251**	-2.950	1.019	182**	-3.251	.888	201**	-3.146	.898	195**	
<i>easy</i> CBM MCRC				4.178	.561	.470**	1.572	.596	.177**	1.521	.600	.171**	
R-CBM							.447	.059	.510**	.400	.082	.456**	
Maze										.277	.328	.074	
$R^2$		.102**			.315**			.483**			.485		
F for change in R <sup>2</sup>		10.250**	:		55.468**	:		58.118**			.712		

## Table 16Linear Regression Results for Seventh Grade CBM on MEAP Reading.

*Note:* FRL = Students qualifying for free and reduced price school lunch. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. \*p < .05. \*\*p < .01. n = 162.

# Table 17Linear Regression Results for Eighth Grade CBM on MEAP Reading.

	Model 1				Model 2			Model 3			Model 4			
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β		
FRL	-5.637	4.239	238**	-7.597	3.578	116	-6.687	3.210	102	-5.598	3.158	085		
<i>easy</i> CBM MCRC				4.965	.477	.566**	3.498	.471	.399**	3.151	.472	.359**		
R-CBM							.339	.045	.396**	.232	.055	.271**		
Maze										.804	.243	.215**		
$R^2$		.057**			.362**			.489**			.513**			
F for change in R <sup>2</sup>		13.608**	¢		108.167**	¢		56.185**			10.981**			

*Note:* FRL = Students qualifying for free and reduced price school lunch. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. \*p < .05. \*\*p < .01. n = 221.

#### Table 18

	Model 1			Model 2				Model 3			Model 4				Model 5		
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β		
FRL	-1.546	.506	206	294	.387	039	290	.387	039	279	.391	037	254	.391	034		
MEAP-Y1				.086	.007	.680**	.085	.007	.673**	.085	.007	.671**	.083	.007	.652**		
ODR							032	.043	038	030	.044	036	001	.050	001		
Total Absences Course Failure										.000	.002	013	.000 120	.002 .096	.012 083		
$\mathbb{R}^2$		.043**			.477**			.478			.478			.482			
F for change in R <sup>2</sup>		9.342**	k		173.344**	:		.563			.058			1.549			

*Linear Regression Results for Eighth Grade EWS on ACT Explore*<sup>®</sup> *Reading.* 

*Note:* MEAP-Y1 = Michigan Education Assessment Program reading subtest for previous year. FRL = Students qualifying for free or reduced price lunch. ODR = Total Office Discipline Referrals. Total Absences = Total number of period absences. Course Failure = Total number of courses failed \*p < .05. \*\*p < .01. n = 180.

## Table 19Linear Regression Results for Eighth Grade CBM on ACT Explore<sup>®</sup> Reading.

	Model 1				Model 2	2		Model 3		Model 4			
Variable	В	SE B	β	В	SE B	β	В	SE B	β	В	SE B	β	
FRL	-1.370	.502	183**	533	.441	071	493	.416	066	335	.397	045	
Ethnicity	229	.123	124	238	.106	129**	184	.100	100	130	.096	070	
<i>easy</i> CBM MCRC				.512	.058	.513**	.380	.060	.380**	.314	.059	.315**	
R-CBM							.031	.006	.314**	0.11	.007	.116	
Maze										.148	.030	.347**	
$R^2$		.057**			.308**			.387**			.448**		
F for change in R <sup>2</sup>		6.644**			78.604**	<		27.920**			23.829**		

*Note:* FRL = Students qualifying for free and reduced price school lunch. MCRC = Multiple Choice Reading Comprehension R-CBM = Reading-Curriculum Based Measure. Maze = Maze Reading Comprehension. \*p < .05. \*\*p < .01. n = 227.

	ELL	SpEd	FRL	White	Black	Hispanic	2 or More Races	Asian	Reading	Math
School	3%	14%	53%	41%	25%	18%	12%	5%	65%	37%
State*	4%	13%	47%	69%	18%	6%	2%	3%	73%	35%
National**	9%	13%	48%	52%	16%	24%	3%	5%	***36%	***35 %

Table 20Comparison Data for Enrollment and Academic Achievement for 2012.

\* State enrollment and achievement data for the State of Michigan provided by the Michigan Department of Education mischooldate.org; \*\*National enrollment data provided by National Center for Education Statistics nces.ed.gov. School and state reading and math data percentage of students proficient for MEAP for 8<sup>th</sup> grade. National reading and math achievement data percentage proficient using National Assessment of Educational Progress, nces.ed.gov/nationsreportcard/naepdata

BIBLIOGRAPHY

#### BIBLIOGRAPHY

- ACT Research and Policy. (2013). *What are the ACT college readiness benchmarks?* Retrieved: http://www.act.org/research/policymakers/pdf/benchmarks.pdf
- Allensworth, E. M., & Easton, J. Q. (2005). The on-track indicator as a predictor of high school graduation. Chicago: Consortium on Chicago School Research.
- Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., ... & Williams, K. L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review*, 33, 218-233.
- Baker, D. L., Biancarosa, G., Park, B. J., Bousselot, T., Smith, J. L., Baker, S. K., ... & Tindal, G. (2014). Validity of CBM measures of oral reading fluency and reading comprehension on high-stakes reading assessments in Grades 7 and 8. *Reading and Writing*, 1-48. doi: 10.1007/s11145-014-9505-4
- Balfanz, R. (2009). Putting middle grades students on the graduation path. *National Middle School Association*. Retrieved June, 1, 2010.
- Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235. doi:10.1080/00461520701621079
- Barry, T. D., Lyman, R. D., & Klinger, L. G. (2002). Academic underachievement and attentiondeficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology*, 40(3), 259-283 doi:10.1016/s0022-4405(02)00100-0
- Barth, A. E., Stuebing, K. K., Fletcher, J. M., Cirino, P. T., Romain, M., Francis, D., & Vaughn, S. (2012). Reliability and validity of oral reading fluency median and mean scores among middle grade readers when using equated texts. *Reading psychology*, 33(1-2), 133-161. doi:10.1080/02702711.2012.631863
- Baydar, N., Brooks-Gunn, J., & Furstenberg, F. F. (1993). Early warning signs of functional illiteracy: Predictors in childhood and adolescence. *Child development*, 64(3), 815-829. doi:10.2307/1131220
- Calhoon, M. B., & Petscher, Y. (2013). Individual and group sensitivity to remedial reading program design: Examining reading gains across three middle school reading projects. *Reading and Writing*, 26(4), 565-592. doi:http://dx.doi.org/10.1007/s11145-013-9426-7

- Casillas, A., Robbins, S., Allen, J., Kuo, Y. L., Hanson, M. A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, 104(2), 407. doi:10.1037/a0027180
- Cohen, J. S., & Smerdon, B. A. (2009). Tightening the dropout tourniquet: Easing the transition from middle to high school. *Preventing School Failure: Alternative Education for Children and Youth*, 53(3), 177-184. doi:10.3200/psfl.53.3.177-184
- Cummings, K. D., Biancarosa, G., Schaper, A., & Reed, D. K. (2014). Examiner error in curriculum-based measurement of oral reading. Journal of School Psychology, 52(4), 361–375. doi:10.1016/j.jsp.2014.05.007
- Deboy, S. L. (2013). The Predictive Relationship Between Oral Reading Fluency and Comprehension As It Relates to Minority Students. Retrieved from: https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/13294/Deboy\_oregon\_01 71A 10726.pdf?sequence=1
- Deno, S. L. (1985). Curriculum-based measurement: the emerging alternative. *Exceptional children*. 52(3), 219-232. Retrieved from: http://psycnet.apa.org/psycinfo/1986-10522-001
- Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure: Alternative Education for Children and Youth*, *36*(2), 5-10. doi:10.1080/1045988x.1992.9944262
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*(3), 184-192. Retrieved from: http://files.eric.ed.gov/fulltext/EJ785942.pdf
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30(4), 507-524. Retrieved from: http://mdestream.mde.k12.ms.us/sped/toolkit/articles/Assessment/Deno%20Sch%20Psyc h%20Rev%202001%20CBM%20overview.pdf
- Denton, C. A. (2012). Response to Intervention for Reading Difficulties in the Primary Grades: Some Answers and Lingering Questions. *Journal of Learning Disabilities*, 45(3), 232– 243. doi:10.1177/0022219412442155
- Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., ... Francis, D. J. (2011). The Relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading : The Official Journal of the Society for the Scientific Study of Reading*, 15(2), 109–135. doi:10.1080/10888431003623546

- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A Synthesis of Reading Interventions and Effects on Reading Comprehension Outcomes for Older Struggling Readers. *Review of Educational Research*, 79(1), 262–300. doi:10.3102/0034654308325998
- Espin, C., Wallace, T., Lembke, E., Campbell, H. and Long, J. D. (2010), Creating a Progress-Monitoring System in Reading for Middle-School Students: Tracking Progress Toward Meeting High-Stakes Standards. Learning Disabilities Research & Practice, 25: 60–75. doi: 10.1111/j.1540-5826.2010.00304.x
- Fuchs, D., & Fuchs, L.S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 92-99. doi:10.1598/RRQ.41.1.4
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RtI: A next-generation approach to multilevel prevention. *Exceptional Children*, 78, 263–279. doi:10.1177/001440291207800301
- Fuchs, D., Fuchs, L. S., & Compton, D.L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly*, 27(4), 216-228. doi:10.2307/1593674
- Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children*, 76(3), 301-323. doi:10.1177/001440291007600304
- Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly*, 25, 33-45. doi:10.2307/1511189
- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention for the learning disabilities construct. *Learning Disabilities Research & Practice*, 18(3), 157-171. doi:10.1111/1540-5826.00072
- Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489. doi:10.5812/ijem.3505
- Goffreda, C. T., Diperna, J. C. and Pedersen, J. A. (2009), Preventive screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Psychol. Schs., 46: 539–552. doi: 10.1002/pits.20396
- Hemphill, F. C., & Vanneman, A. (2011). Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2011-459. National Center for Education Statistics. doi:10.1037/e595292011-001

- Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636-644. doi:10.1598/RT.59.7.3
- Heppen, J. B., & Therriault, S. B. (2008). Developing Early Warning Systems to Identify Potential High School Dropouts. Issue Brief. *National High School Center*.
- Hernandez, D. J. (2011). Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation. *Annie E. Casey Foundation*.
- Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychological bulletin*, *111*(1), 127. doi:10.1037/0033-2909.111.1.127
- Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review*, 40(1), 108.
- IDEA (2004). Individuals with Disabilities Education Improvement Act. Public Law 108-446. doi:10.1007/springerreference\_69963
- Irvin, P. S., Alonzo, J., Lai, C. F., Park, B. J., & Tindal, G. (2012). Analyzing the Reliability of the easyCBM Reading Comprehension Measures: Grade 7. Technical Report# 1206. Behavioral Research and Teaching.
- Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. *Identification of learning disabilities: Research to practice*, 99-149
- Jerald, C. D. (2006). Identifying Potential Dropouts: Key Lessons for Building an Early Warning Data System. A Dual Agenda of High Standards and High Graduation Rates. *Achieve*, *Inc*.
- Katsiyannis, A., Zhang, D., Ryan, J. B., & Jones, J. (2007). High-stakes testing and students with disabilities: Challenges and promises. *Journal of Disability Policy Studies*, 18(3), 160-167. doi:10.1177/10442073070180030401
- Kennelly, L., & Monrad, M. (2007). Approaches to Dropout Prevention: Heeding Early Warning Signs With Appropriate Interventions. doi:10.1037/e538292012-001
- Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of school psychology*, 52(4), 377-405. doi:10.1016/j.jsp.2014.06.002

- Kratochwill, T. R., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models: Implications for Response to Intervention. *School Psychology Review*, 36(4), 618-631.
- McDermott, R., Raley, J. D., & Seyer-Ochi, I. (2009). Race and class in a culture of risk. *Review* of *Research in Education*. 33, 101-116. Accessed; http://www.jstor.org/stable/40588119
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*, 193-203.
- McIntosh, K., Goodman, S., & Bohanon, H. (2010). Toward True Integration of Academic and Behavior Response to Intervention Systems: Part One--Tier 1 Support. *Communiqué*, 39(2), 1-14.
- Michigan Department of Education. (2012). Michigan Educational Assessment Program: Technical Report 2011-2012. Retrieved from http://www.michigan.gov/documents/mde/MEAP\_2010-2011\_Technical\_Report\_394693\_7.pdf
- National Center on Response to Intervention (2012). RtI state database. Retrieved from http://state.rti4success.org/
- National Research Center on Learning Disabilities (2006). *A tiered service-delivery model*. Retrieved from http://www.nrcld.org/rti\_manual/pages/RTIManualSection3.pdf
- NCLB (2002). No Child Left Behind Act. Public Law 107-115. doi:10.1007/springerreference\_223926
- Neild, R. C., Balfanz, R., & Herzog, L. (2007). An early warning system. *Educational leadership*, 65(2), 28-33.
- Nigg, J. T., Hinshaw, S. P., Carte, E. T., & Treuting, J. J. (1998). Neuropsychological correlates of childhood attention-deficit/hyperactivity disorder: Explainable by comorbid disruptive behavior or reading problems? Journal of Abnormal Psychology, 107(3), 468–480. doi:10.1037/0021-843x.107.3.468
- O'Connor, C., Hill, L. D., & Robinson, S. R. (2009). Who's at risk in school and what's race got to do with it?. *Review of research in education*, *33*(1), 1-34.
- Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, 15(12), 1-9.
- Park, B. J., Alonzo, J., & Tindal, G. (2011). The Development and Technical Adequacy of Seventh-Grade Reading Comprehension Measures in a Progress Monitoring Assessment System. Technical Report# 1102. *Behavioral Research and Teaching*.

- Peng, C. Y. J., & So, T. S. H. (2002). Logistic regression analysis and reporting: A primer. Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1(1), 31-70. doi:10.1207/s15328031us0101\_04
- Pinkus, L. (2008). Using early-warning data to improve graduation rates: Closing cracks in the education system. *Washington, DC: Alliance for Excellent Education*.
- Reed, D. K., & Sturges, K. M. (2012). An Examination of Assessment Fidelity in the Administration and Interpretation of Reading Tests. Remedial and Special Education, 34(5), 259–268. doi:10.1177/0741932512464580
- Reed, D. K., Wexler, J., & Vaughn, S. (2012). RTI for reading at the secondary level: Recommended literacy practices and remaining questions. Guilford Press.
- Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology*, 36(5), 759-770. doi:10.1007/s10802-007-9208-2
- Roberts, G., Vaughn, S., Fletcher, J., Stuebing, K., & Barth, A. (2013). Effects of a Response-Based, Tiered Framework for Intervening With Struggling Readers in Middle School. *Reading research quarterly*, 48(3), 237-254. doi:10.1002/rrq.47
- Saez, L., Park, B., Nese, J. F., Jamgochian, E., Lai, C. F., Anderson, D., ... & Tindal, G. (2010). Technical Adequacy of the easyCBM Reading Measures (Grades 3-7), 2009-2010 Version. Technical Report# 1005. *Behavioral Research and Teaching*.
- Salvia, J., Ysseldyke, J., & Bolt, S. (2012). Assessment: In special and inclusive education. Cengage Learning.
- Silberglitt, B., & Hintze, J. M. (2007). How Much Growth Can We Expect? A Conditional Analysis of R-CBM Growth Rates by Level of Performance. *Exceptional Children*, 74(1), 71-84. doi:10.1177/001440290707400104
- Smartt, S., & Reschly, D. (2007). Barriers to the preparation of highly qualified teachers in reading. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.ncctq.org/tqbrief.php
- Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36,* 24-33. doi:10.1177/00222194030360010401
- State of Michigan. Department of Education. (2011). Approval of Recommended New Cut Scores on the Michigan Educational Assessment Program (MEAP) and Michigan Merit

Examination (MME) Consistent with Career and College Readiness. by M. Flanagan. Lansing, MI (Memorandum).

- Stevenson, N. (2015). Predicting proficiency on statewide assessments: A Comparison of curriculum-based measures. *Journal of Educational Research*. doi: 10.1080/00220671.2014.910161
- Sugai, G. & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Ingegration of multi-tiered systems approaches. *Exceptionality*, 17, 223-237.
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., ... & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions*,2(3), 131-143. doi:10.1080/09362830903235375
- Suh, S., & Suh, J. (2007). Risk factors and levels of risk for high school dropouts. *Professional School Counseling*, *10*(3), 297-306.
- Sullivan, C. J., Childs, K. K., & O'Connell, D. (2010). Adolescent risk behavior subgroups: An empirical assessment. *Journal of youth and adolescence*, 39(5), 541-562. doi:10.1007/s10964-009-9445-5

Tabachnick, B. G. & Fidell, L. S. (2013). Using multivariate statistics, 6/e. Pearson, Inc.

- Tindal, G., & Nese, J. F. (2011). Applications of curriculum-based measures in making decisions with multiple reference points. *Advances in Learning and Behavioral Disabilities*, *24*, 31-58. doi:10.1108/s0735-004x(2011)0000024004
- Tindal, G., Nese, J. F., & Alonzo, J. (2009). Criterion-Related Evidence Using easyCBM [R] Reading Measures and Student Demographics to Predict State Test Performance in Grades 3-8. Technical Report# 0910. *Behavioral Research and Teaching*.
- Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, 7(1), 40-53. doi:10.1177/106342669900700105
- Tolar, T. D., Barth, A. E., Francis, D. J., Fletcher, J. M., Stuebing, K. K., & Vaughn, S. (2011). Psychometric properties of maze tasks in middle school students. Assessment for Effective Intervention, doi:10.1177/1534508411413913
- Vanderheyden, A. M., & Tilly, W. D. (2010). *Keeping RTI on track: How to identify, repair and prevent mistakes that derail implementation*. LRP Publications.
- Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009). Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the

National Assessment of Educational Progress. Statistical Analysis Report. NCES 2009-455. *National Center for Education Statistics*.

- Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. D., ... Francis, D. J. (2010). Response to Intervention for Middle School Students With Reading Difficulties: Effects of a Primary and Secondary Intervention. *School Psychology Review*, 39(1), 3– 21.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*(3), 137-146. doi:10.1111/1540-5826.00070
- Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., Barth, A., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*, 45(6), 515-525. doi:10.1177/0022219411402692
- Wanzek, J., Vaughn, S., Roberts, G., & Fletcher, J. M. (2011). Efficacy of a reading intervention for middle school students with learning disabilities. *Exceptional Children*, 78(1), 73–87.
- Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*, 26(4), 207-214. doi:10.1177/07419325050260040301
- Yeo, S. (2009). Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education*, 31(6), 412–422. doi:10.1177/0741932508327463