2013

The utility of curriculum-based measurement within a multitiered framework: establishing cut scores as predictors of student performance on the Alaska standards-based assessment

David E. Legg

Pepperdine University

Graduate School of Education and Psychology

THE UTILITY OF CURRICULUM-BASED MEASUREMENT WITHIN A MULTITIERED FRAMEWORK: ESTABLISHING CUT SCORES AS PREDICTORS OF STUDENT PERFORMANCE ON THE ALASKA STANDARDS-BASED ASSESSMENT

A dissertation presented in partial satisfaction

of the requirement for the degree of

Doctor of Education in Educational Leadership, Administration, and Policy

by

David E. Legg

March 2013

Devin Vodicka, Ed.D. – Dissertation Chairperson

This dissertation, written by

David E. Legg

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Doctoral Committee:

Devin Vodicka, Ed.D., Chairperson

Linda Purrington, Ed.D.

Doug Leigh, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

DEDICATION

This dissertation is dedicated to my family for the sacrifices they made allowing me to pursue my doctoral studies. To my loving wife Sheri, thank you for bearing the extra burden of caring for our children when my studies occupied my time and attention. Without your unwavering commitment, the pursuit of my doctoral studies would not have been possible. Thank you! To my three children: Taylor, Hunter, and Paisley, I look forward to recapturing the time lost with each of you over the last 6 years.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude for the years of encouragement, wisdom and support given to me by my committee. Thank you, Dr. Purringtion, for your unwavering commitment that extended from my initial interview through my final defense. Thank you, Dr. Vodicka for agreeing to chair my committee through the final defense, even though your circumstances changed along the way. You truly understood what I was attempting to accomplish and were always able to keep me on track with my analyses. Dr. Leigh, your constructive feedback and guidance regarding research design was highly valued. I sincerely appreciate the time and dedication that each of you has provided over the last few years as I completed my doctoral studies. I would like to extend a special thank you to Dr. Steven Atwater for supporting me and allowing me to conduct this study. Preliminary statistical assistance on statistical procedures was graciously provided by Dr. Ben Ditcowsky.

VITAE

David E. Legg

Education

Doctor of Education, Educational Leadership, Administration, and Policy, Pepperdine University, Malibu, California, 2013

Post Graduate Certificate: Superintendent, University of Alaska, Anchorage, Alaska, 2003

Masters in Educational Leadership, University of Alaska, Anchorage, Alaska, 2000

Bachelor of Arts Special Education, Arizona State University, Phoenix, Arizona, 1997

Recent Professional Experience

2011-Present	Anchorage School District, Anchorage, Alaska *Assistant Principal: Freshman House-Dimond High*
2009-2011	Anchorage School District, Anchorage, Alaska *Special Education Supervisor*
2005-2009	Kenai Peninsula Borough School District, Soldotna, Alaska *Assistant Director of Pupil Services*
2004-2005	Southwest Region Schools, Dillingham, Alaska *Associate Superintendent/Director of Special Education*
2002-2004	Southwest Region Schools, Manokotak, Alaska *Principal of K-12 School*
2000-2002	Southwest Region Schools, Ekwok, Alaska *Principal/Teacher/Special Educator*
1997-2000	North Slope Borough School District, Wainwright, Alaska *Teacher of 7th and 8th Grade Math, English, and Resource*

ABSTRACT

The purpose of this study was to explore the relationship between student performance on Reading Curriculum-based Measures (R-CBM) and student performance on the Alaska's standards based assessment (SBA) administered to students in Studied School District Grade 3 through Grade 5 students in the Studied School District as required by Alaska's accountability system. The 2 research questions were: (a) To what extent, if at all, is there a relationship between student performance on the R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and student performance on the Alaska SBA administered in the spring of the same school year in the SSD? (b) To what extent, if at all, can cut scores be derived for each of the 3 R-CBM testing windows in the fall, winter, and spring that predict success on the Alaska SBA administered in the spring of the same school year in the SSD? The Study School District (SSD) served approximately 9,500 students, with 14% of students eligible for special education services. The enrollment was 81% Caucasian, 10% Alaska Native, 3% Hispanic, 3% multiethnic, and 4% as the total of American Indian, Asian, Black, and Native Hawaiian/Pacific Islander. The sample was 3rd (n = 472), 4th (n = 435), and 5th (n = 517) graders and consisted of all students with an Alaska SBA score and an R-CBM score for each of the 3 administrations of the R-CBM used in the 2009-2010 (FY10) and 2010-2011 (FY11) years. Pearson correlations were significant between R-CBM scores across 3rd, 4th, and 5th grades and the same grade Alaska SBA scores for FY10 data, $r = .689$ to $r = .728$, $p < .01$. A test of the full model with R-CBM as predictor against a constant-only model was statistically reliable, $p < .001$. The R-CBM reliably distinguished between passing and failing the Alaska SBA for students in Grades 3 through 5. Criterion validity of the

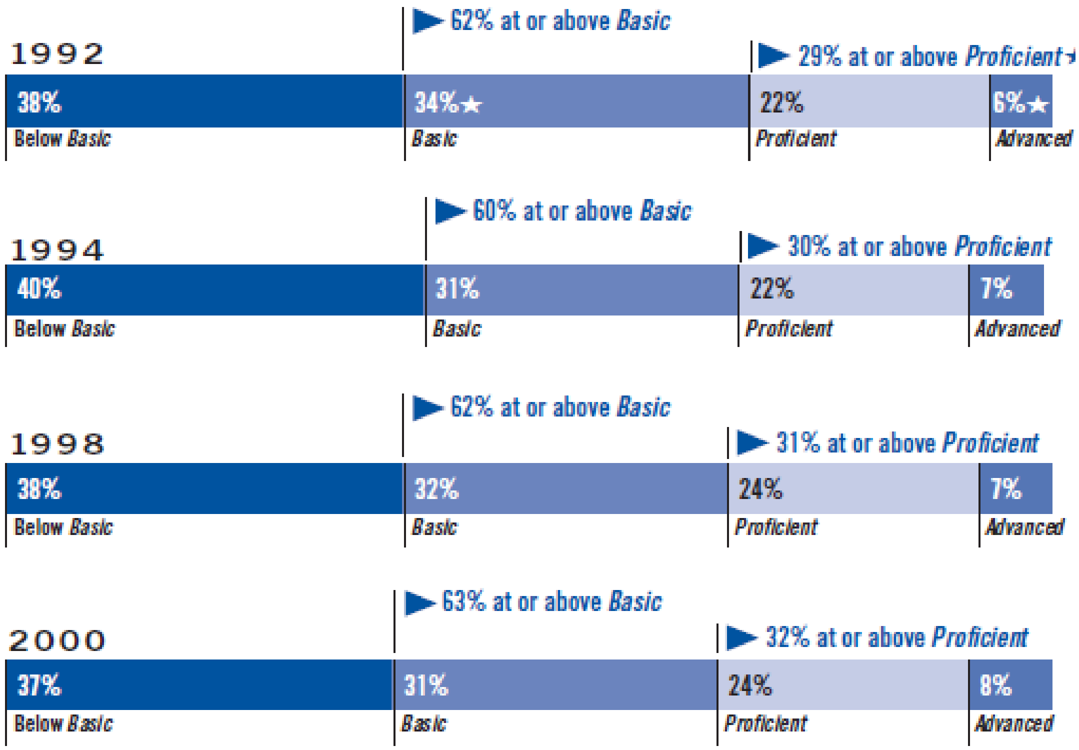
cut scores was ascertained by applying scores to the FY11 data and yielded adequate levels of sensitivity from 49% to 88% while specificity levels ranged from 89% to 97%.

Chapter 1: Introduction

Background

In the recent report *Forces Changing Our Nation's Future* prepared for the National Commission on Adult Literacy, Kirsch, Braun, Yamamoto, and Sum (2007) highlighted what many educators instinctively know: Students' levels of educational attainment and earning potential are tied to literacy. Similarly, Brigman, Webb, and Campbell (2007) studied low student achievement in reading and a variety of social problems, including dropout rates, delinquency, and teen pregnancy and reported significant correlations between all of these variables. Although literacy has been acknowledged as a requisite skill known to impact significantly both academic and economic success, the educational system continues allowing too many students to advance through school without acquiring basic literacy skills. According to the National Center for Education Statistics (NCES, 2003), 37% of students enter the fourth grade with below grade level basic reading skills. This NCES figure is especially alarming since remediation has not proven to be effective (Griffiths, Parson, Burns, VanDerHeyden, & Tilly, 2007). See Figure 1.

The *No Child Left Behind Act of 2001* (NCLB, 2002) was passed in an effort to increase literacy by requiring all states to put measures in place to ensure that students are proficient in reading by the third grade. In order for all children to be proficient readers by the third grade, traditional methods of intervention including monitoring student progress, evaluating the effectiveness of instruction through formative assessments, and identifying students with learning disabilities have been replaced with different monitoring techniques and early identification of at-risk status as part of the Individuals with Disabilities Education Act (IDEA, 1997) reauthorization in 2004.



★ Significantly different from 2000.
NOTE: Percentages within each reading achievement level range may not add to 100, or to the exact percentages at or above achievement levels, due to rounding.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2000 Reading Assessments.

Figure 1. Percentage of fourth-grade students at or above reading achievement levels and within each achievement level range: 1992-2000. Reproduced with permission.

Historically, children who experience difficulties in reading have often not been identified as having learning disabilities until the fourth or fifth grade (Ofiesh, 2006). The delay in the identification of children with learning disabilities has been blamed on the identification process itself. Until the IDEA reauthorization in 2004, children were identified as learning disabled using the discrepancy model (Ofiesh, 2006). This model required learning disabled students to have a severe discrepancy between intellectual ability and academic achievement. A severe discrepancy usually required exact estimates between student performance, as measured by scores from both an achievement test and an intelligence test. With the discrepancy model, commonly referred to as the *wait to fail* model (Berkeley, Bender, Peaster, & Saunders, 2009), students are generally not identified as having a learning disability prior to the third grade and often are not identified until the fifth grade (Ofiesh, 2006). Because of the delay in the prevention, identification, and intervention of reading difficulties, children are not likely to receive services when they can benefit from them the most prior to the third grade.

Recognizing the need for early identification and intervention for at-risk children, the Office of Special Education Programs (OSEP) began the learning disabilities initiative (Bradley & Danielson, 2004). The purpose of the learning disabilities initiative was to examine the key issues surrounding the determination of learning disabilities and to use the findings to assist in future policy decisions. As a result of this initiative, an alternative way for determining learning disabilities became needed. Moreover, consensus among participating researchers regarding the use of Response to Intervention (RTI) in the determination of learning disabilities showed the following:

> There should be alternative ways to identify individuals with [learning disabilities] in addition to achievement testing, history, and observations of the child. Response to scientifically valid and generally effective intervention is the most promising method of alternative identification and can both promote effective practices in the schools and help to close the gap between identification and treatment. (Bradley & Danielson, 2004, p. 188)

Increased accountability. IDEA originally guaranteed students with disabilities access to the general curriculum. The 2004 Individuals with Disabilities Education Improvement Act (IDEIA) reauthorized IDEA and emphasized results and accountability. Similarly, implementation of NCLB has placed more accountability on schools to increase student achievement by reaching established grade-level benchmarks on standardized assessments. Hence, failure of schools to make *adequate yearly progress* (AYP) with students in need can lead to government-sanctioned consequences including replacement of school staff (NCLB, 2001). The authors of IDEIA used the learning disabilities initiative and incorporated facets of that initiative into IDEIA. IDEIA has not eliminated the discrepancy model for learning disabilities altogether but has a provision that requires the states' departments of education to allow individual school districts the option of using RTI to identify learning disabilities that students might have.

Response to intervention. The provision in IDEIA (2004) allowing schools to use RTI and the increased accountability called for by NCLB (2002) for student achievement by requiring schools to make AYP have led to a heightened interest in RTI by schools (Buffum, Mattos, & Webber, 2009). RTI relies on the use of tiered interventions for students who are not making sufficient progress in the general curriculum (IDEIA, 2004). With a shift toward intervention rather than remediation, assessments to identify struggling readers prior to the third grade need to be available. Good, Simmons, and Kame'enui (2001) reported the key elements of a prevention-oriented approach as the "ability to predict reading success and difficulty early and to inform instruction responsively" (p. 260). In order to properly identify students and provide appropriate interventions, efficient methods of monitoring students' general reading ability are needed. According to Good et al., schools should adhere to the following principles when considering both the design and the use of assessments:

1. Intervene early and strategically during critical windows of reading development.
2. Develop and promote a comprehensive system of instruction based on a research-based core curriculum and enhancement programs.
3. Use and rely on formative, dynamic indicators of student performance to identify need, allocate resources, and design and modify instruction.

4. Address reading failure and reading success from a schoolwide systemic perspective. (p. 260)

Curriculum-based Measurement (CBM) directly or indirectly addresses the need for the previously mentioned principles for the design of assessment and show promise for use with RTI (Silberglitt, Burns, Madyun, & Lail, 2006).

CBMs. The CBM was developed in 1977 as a method of monitoring the effectiveness of special education instruction (Deno, S., 2003). Data from CBM monitoring can be graphed and analyzed for decision making. CBM uses brief, timed measures of students' general ability (Shinn, 2007). The body of research around CBM continues to grow and has shown CBM to have significant value to educators in decision making (Deno, S., 2003; Deno, Fuchs, Marston, & Shin, 2001). With greater demands for schools to show increasing student performance on high-stakes exams, school officials are searching for more timely and cost-efficient ways to identify struggling readers and provide appropriate interventions in their efforts to ensure that all students are proficient readers by the third grade. Research has demonstrated multiple uses for CBM, including the improvement of instructional programs, predicting student performance against specific criteria, developing grade-level norms, and as a universal screener for identifying students at risk of failing (Deno, S., 2003). Crawford, Tindal, and Stieber (2001); S. Deno (2003); and Fuchs and Fuchs (2004) revealed that CBMs can accurately predict student performance on high-stakes tests. In recent years, several studies have been conducted specifically for examining the value of CBMs and their utility in predicting student success on state standardized high-stakes assessments (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good et al., 2001; McGlinchey & Hixson, 2004; Wood, 2006). Since many states begin high-stakes testing in the third grade, much of the emphasis for existing studies has been directed at first through third grades (Good et al., 2001).

The requirement that all children should be fluent readers by third grade coupled with the requirements of NCLB has led states to implement high-stakes testing beginning in the third grade. With mounting pressure for schools to prevent reading difficulties through early identification and prevention, affordable assessments need to be available that can efficiently and accurately identify struggling readers prior to the third grade. Teachers and administrators are faced with the challenge of identifying formative assessments to assist in measuring whether students are on track to meet desired benchmarks or are at risk of failing to meet determined benchmarks. While several recent studies have included CBMs to predict student success on standardized assessments (Good et al., 2001; McGlinchey & Hixson, 2004), studies of a longer duration continue to be needed to make these determinations. Additional studies of different populations are needed to assist in determining the degree to which the findings can be generalized between different populations.

Although one of the original purposes of CBM was to assist teachers in the identifying students whose performance was discrepant from classroom peers and in need of a diagnostic assessment (Deno, S., 1985, p. 230), a growing body of research has indicated CBM to be a reliable and valid predictor of student achievement on standardized assessments. Studies have been conducted in several states for setting benchmarks for fluency in reading through the third grade (Sibley, Biwer, & Hesch, 2001). CBMs can be used for universal screening, a process accepted as a requisite component to implementing RTI. In addition to the continued research needed on the utility of using the CBM to aid in instructional decisions, longitudinal research has been called for to assist in determining the ability to predict student achievement in the higher grades (Good et al., 2001).

Although the CBM has been known to have certain utility for more than 25 years, it appears that most teachers and administrators continue not to be fully aware of the utility of CBM use or how to use it to assist with identifying at-risk students and with ongoing progress monitoring. In addition to the value of using a CBM to identify at-risk students, the CBM can be used frequently with reliability for regular progress monitoring of students once they have been placed in tiered interventions or in special education. The resulting data provide teachers with regular feedback and alert them regarding the efficacy of a specific intervention. The use of the CBM has several advantages over norm-referenced assessments.

CBMs are short, 1-minute timed measures designed to represent student performance as related to the curriculum. CBMs are inexpensive and may be given frequently. “Because the progress-monitoring component of RTI should employ tools that are scientifically based, it seems logical that CBM would be the primary tool in the RTI process when there are concerns about a student’s basic skills” (Shinn, 2007, p. 608). This study examined the predictive value of a curriculum-based measure of oral reading fluency (R-CBM) to the Alaska SBA.

Statement of the Problem

The relationship between the number of words read correct (WRC) on R-CBM and the Alaska Standard Based Assessment (SBA) was examined in this study. The R-CBM can be used as a tool in many instructional decisions (e.g., predicating performance on criteria, developing student norms, cut scores, and screening for students at risk of academic failure). S. Deno (1985) suggested using normative data from a local sample to determine how students preform against their peers in the same curriculum. Hasbrouck and Tindal (1992) argued that in order for norms to be useful in making instructional decisions that larger samples representative of both student ability and geographic location are more meaningful. Norms allow educators a benchmark against student proficiency. Unfortunately, no single universally accepted definition of proficiency has been developed. Mandates from NCLB have led to integrated assessments that provide state education agencies (SEAs) and local education agencies (LEAs) as constructs for demonstrating growth against adopted standards. Researchers have begun looking to CBM as a predictor variable when monitoring both district and student progress towards AYP. This approach if adopted can provide educators with a specific definition of proficiency (Silberflitt & Hintze, 2005).

Emerging research in several states now shows that reading student performance on R-CBM has promise for predicting student success on high-stakes testing (Kovaleski, 2007; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). Studies have been done in several states showing a strong correlation between R-CBM fluency and state high-stakes assessments. However, no study has been done in the state of Alaska. Alaska has developed its own assessment and needs to establish a correlation between WRC on commercially available R-CBM and the Alaska SBA. Without exploring the relationship between student performance on R-CBM tools and student performance on the Alaska SBA, Alaska must rely on benchmarks and or norms established by other states to predict student performance as well as normative data. No national norms have been established because each state establishes its own procedures, norms, and expectations for benchmarking and assessment (Shin, 2007). The lack of national norms poses a problem for ensuring all students, regardless of where they attend school, receive the same quality and level of education.

When forced to rely on norms developed in other states, districts outside of the norm group are at a disadvantage in that comparisons made to other students are not representative as a measure of the same curriculum. A key component of RTI is that core curriculum must be effective for 80% of students. This determination cannot be made using norms based on performance on other state exams. Local norms and cut scores are needed to effectively implement RTI and ensure that student are making adequate progress toward mastery of local standards and thus are likely to pass grade-level high-stakes exams. Previously established benchmarks might not share the same correlation with Alaska tests as have been established in other states using their assessments.

Purpose of the Study

The purpose of this study was to explore the relationship between student performance on R-CBMs and student performance on the Alaska's SBA administered to students in Studied School District Grade 3 through Grade 5 students in the Studied School District as required by Alaska's accountability system. Secondly, if a statistically significant relationship between student performance on R-CBM tools and Alaska SBA was observed, the researcher would examine the efficacy of deriving cut scores via logistic regression for use in predicting whether students are on track to meet proficiency requirements on the Alaska SBA. Currently, Alaska is required to administer a criterion-referenced SBA to all students grade three through ten. This study builds on existing studies in two key areas. Cross validation of the cut scores established in this study were applied to the next year's student performance data on the Alaska SBA for the purpose of determining their ability to properly predict students' performance on the spring administration of the Alaska SBA. A secondary purpose of this study was to expand findings from previous studies to additional populations and other assessments.

Research Questions

Two broad research questions framed this study across Grades 3, 4, and 5. Within each grade, the research questions were applied to interval data obtained through the triennial administration of R-CBM. The research questions were:

1. To what extent, if at all, is there a relationship between student performance on R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and student performance the Alaska SBA administered in the spring of the same school year in the SSD?
2. To what extent, if at all, can cut scores be derived for each of the three R-CBM testing windows in the fall, winter, and spring that predict success on the Alaska SBA administered in the spring of the same school year in the SSD?

Operational Definitions of Variables

Adequate yearly progress (AYP). AYP is an individual state's measure of progress toward the goal of 100% of students meeting state academic standards in at least reading/language arts and math (NCLB, 2001). The Alaska SBA is used to meet AYP in Alaska.

Curriculum-based measure (CBM). "An approach to assessing student growth in basic skills through frequent assessments" (Deno, S., 2003). The R-CBM was used in this study as an independent variable.

Curriculum-based measure of reading (R-CBM). "A standardized, individually administered test of accuracy and fluency with connected text" (University of Oregon, n.d., para. 1) was for the R-CBM in this study. The R-CBM is a "1 minute standardized measure of oral reading of graded passages to administer for individual students" (AIMSweb, 2011b, column 4, para. 1).

High-stakes reading assessment. A reading "test [is] used to provide results that have important, direct consequences for examinees, programs, or institutions involved in the testing" (Alaska Department of Education & Early Development [ADEED], 2011a, p. 30). Third, fourth and fifth-grade criterion-referenced reading required by the state of Alaska is tested using the Alaska SBA to fulfill the testing requirements of the NCLB and was used to measure this variable.

Progress monitoring. "A scientifically based practice that is used to assess students' academic performance and evaluate the effectiveness of instruction. Progress monitoring can be implemented with individual students or an entire class" (National Center on Student Progress Monitoring, n.d., What is Progress Monitoring, para. 1). The R-CBM represented the application of progress monitoring in this study.

Response to intervention (RTI). "RTI is the practice of providing high-quality instruction and intervention matched to student need, monitoring progress frequently to make decisions about change in instruction or goals and applying child response data to important educational decisions" (Griffiths et al., 2007). The R-CBM represented the method for identifying students for RTI in this study.

Importance of the Study

There have been several studies conducted in other states demonstrating a high correlation between high-stakes standardized assessment and R-CBM. If similar correlations can be established between the Alaska SBA and R-CBM, the findings are expected to support a growing body of similar research. Each of the states in the previous studies has different high-stakes assessments, but the findings of this study further verified findings in previous studies that identified high correlations between R-CBM and standardized assessments. This suggested that the use of R-CBM has evidence of validity. This study differed from previous studies in that it attempted to set statistically derived cut scores rather than relying on normative scores established in previous studies. In doing so, the study offered evidence of convergent validity for using the fall, winter, and spring R-CBM administrations alongside the spring administrations of the Alaska SBA. Moreover, these cut scores were cross validated against the other years of data to determine their ability to classify outcomes on the Alaska SBA across years other than the original sample.

This study identified statistically significant correlations between R-CBM and high-stakes assessments. Schools could use this study's findings to develop local norms as well as statistically derived cut scores and benchmarks as predictors of success on high-stakes exams. Moreover, since R-CBMs are readily available in multiple forms, they can be used for frequent progress monitoring of students. Because R-CBMs are sensitive to small changes in student performance, schools have the ability to make changes in a student's curriculum when the benchmarks indicate a student to be at risk of failing a high-stakes exam.

Assumptions

R-CBM probes are designed to be given using a standard protocol. Teachers responsible for administering the probes were trained in the administration of the probes at the beginning of the school year. Because the investigator was not able to directly observe the administration of the probes, it is assumed that teachers adhered to the standard protocols during the R-CBM and Alaska SBA testing.

Limitations

This study was intended to provide cut scores specific to SSD based on local demographic and population. Due to demographic diversity in the state of Alaska, there is a large percentage of English Language Learners that is not fully represented by the findings of this study. ADEED (2011a) reports that statewide, 54% of the population is Caucasian. The student population in this study was 83% Caucasian and therefore not representative of the demographic makeup of most Alaska school districts. Even though the Alaska SBA is a statewide assessment (ADEED, 2005), the ability to generalize the findings of this study to the rest of the state is not known due to lack of representation of English Language Learners as well as Native Alaska populations that are prevalent in most of the rural Alaska school districts.

Findings of this study were expected to substantiate previous research supporting the use of CBM to predict student achievement on state high-stakes testing (Crawford, et al., 2001; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Sibley et al., 2001; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001) and extended the research to the fifth grade. However, the duration of the data analyzed was delimited to two years, limiting the ability to examine predictability of student achievement more than one year in advance. The findings were not generalizable to other Alaska school districts that might have populations with greater diversity in their demographic makeup. Because this study did not mirror the state's demographic makeup, the findings of this study cannot be generalized to other school districts throughout Alaska.

Delimitations

The requirements of NCLB measure student growth and subsequently label schools effectiveness based on student outcomes on mandated tests. R-CBM norms and benchmarks have been established in multiple states that are strongly tied to state exams. Since each state has their own assessment system, benchmarks need to be established that accurately predict success or failure on states exam. For this reason, this study was limited to Alaska. The SSD of this study is not reflective of the demographic found throughout much of the state and currently relies on national norms as a part of their RTI implementation. Since SSD relies on R-CBM within an RTI framework, cut scores or bench marks need to be established based on their population. These cut scores can in turn be utilized as a measure of the effectiveness of instruction throughout the district. The scores from the districts home school program and school with assessment calendars that did not conform with the administration of the R-CBM were excluded from this analysis. Scores from the district sponsored home school program were excluded because they would not be reflective of instruction by SSD. The amount of instruction between administrations of R-CBM would vary between schools with differing calendars thus potentially effecting the correlations between the ASBA and R-CBM.

Chapter 2: Review of the Literature

High-stakes Assessments

Chapter 1 addressed that the No Child Left Behind Act of 2001 (NCLB, 2002) requires that all students be proficient readers by the third grade. Moreover, by the year 2013, schools are required to ensure that all students score proficient on approved standardized tests. Because of the consequences associated with schools that are not able to meet these stringent requirements, both teachers and administrators are looking for better methods to identify early on those students who are not likely to pass high-stakes assessments.

Alaska Standards-based Assessment

Alaska is second only to Hawaii regarding gaining statehood. Attaining statehood in 1959, Alaska has been a state for a little more than 50 years. Although it is more than twice the size of Texas, covering over 586,000 square miles, the state is comprised of only 53 school districts. School districts in Alaska are generally classified as urban or rural. The larger urban districts are governed by either boroughs or municipalities. Although the school districts have autonomy in selecting superintendents and in district operations, they are accountable to local governments regarding the use of capital and operating budgets (McBeath, Reyes, & Ehrlander, 2008). In addition to state funding, statutes establish and define the local effort that local governments must contribute. Several rural school districts reside in first-class cities and likewise operate as independent school districts. Similarly, there are several borough school districts located in rural settings. However, most rural school districts are governed by one of 19 Regional Educational Attendance Area (REAA) school boards. "These legislativelycreated school districts—products of the rural school decentralization act of 1975—are autonomous. They report directly to the Alaska Department of Education and Early Development (ADEED) without the interference of local government bodies" (McBeath et al., 2008, p. 259).

History of Alaska's Accountability System

Alaska's current accountability system is the result of 15 years of school reform efforts. In 1991, Governor Walter J. Hickle directed the commissioner of education to develop a statewide system of school reform. In November of the same year, a commission of prominent Alaskans was appointed by the governor to identify areas of educational concern. The recommendations of the commission were approved by ADEED in October of 1991, thus becoming the focus of statewide school reform through 1995. This movement for school reform soon became known as Alaska 2000 (AK2K). Over the next 4 years, the work of 10 committees led to new regulations regarding the minimum knowledge students should have across 10 key subject areas. In February 1993, the Standards and Oversight Committee began the work of developing specific content standards in each of the 10 core subject areas. After nearly two years and two periods of public comment, standards for math, science, and English/language arts were adopted and placed into regulation in January of 1995. Under the direction of newly elected Governor Tony Knowles, a new school board headed by Commissioner Shirley Holloway continued with the AK2K initiative by adopting additional standards for world languages and technology. These regulations took effect on March 13, 1995 (ADEED, 2007).

AK2K was soon replaced by a new initiative that continued to build on the principles of continued school improvement. The Quality Schools Initiative (QSI) was comprised of four key areas: high student academic standards and assessments; quality professional standards; family, school, business, and community network; and school excellence standards. This plan evolved as a result of an education summit in the fall of 1996. In addition to the key areas identified at the fall 1996 summit, Commissioner Holloway committed to developing an outcome-based system of education based on specific standards. In keeping the commitment to create an educational system based on standards, legislation was introduced in the spring of 1997 to require all students to pass an exit exam prior to being awarded a high school diploma beginning January 1, 2002 (ADEED, 2007).

Alaska Senate Bill 36, known for introducing drastic changes to the education funding formula, also included several components for QSI. The bill was signed into law in the summer of 1998. Specific to QSI was a mandate for ADEED to adopt performance standards aligned to the Alaska Benchmark Exam. Further, beginning in March of 2000, ADEED mandated that the Alaska Benchmark was to be administered to students in the third, sixth, and eighth grades. The intent of this exam was to determine whether students were meeting established performance standards. Based on the mandate, ADEED quickly established performance standards, which were adopted into regulation in January of 1999. Soon afterward and equally important was the initial field-testing of the mandated assessment, which was to be aligned to the performance standards. By September of 1999, the field-testing of the benchmark assessment was complete and the assessment was adopted into regulation. By March of the following year, the high school exit exam was administered, and a few months later, and prior to the end of 2000, cut scores for proficiency were established for the benchmark exam (ADEED, 2007).

By July of 2001, the state was ready to release the first data regarding student performance on the mandated high school exit exam. The data showed significant disparities between ethnic groups, specifically with particularly low scores for Alaska Native students. Because many of Alaska's indigenous people live in rural settings often not connected by the state's road system, the subpar performance by disproportionate numbers of Alaska Native students is often referred to as the rural/urban divide (McBeath et al., 2008; United States Commission on Civil Rights, 2002). Commissioner Holloway released a memorandum to policymakers and educators throughout the state calling them to question the results and take action to close the achievement gap:

> The data I am releasing today will cause deep soul searching in Alaska. The analysis shows a deep divide in student achievement among ethnic groups. White students score higher than other ethnic groups, much higher on average than Native students. Why is this so? What steps do we need to take to shrink this divide? It's time for debate. It's time to find out. It's time for action. As we more deeply analyze the data, a picture begins to come into focus. It is important that we share this picture with others. By doing so we encourage broader understanding of what the exam results mean, and stimulate debate over what we need to do to improve achievement. It is vital that our data-driven debate be free of political and personal agendas and is focused on students. (ADEED, 2007, p. 10)

In September of 2001, the vision of a comprehensive assessment program was completed with the adoption of the TerraNova, a norm-referenced assessment, to Grades 5 and 9. With this addition, students were tested with the TerraNova in Grades 4, 5, 7, and 9. The Alaska Benchmark was administered to students in Grades 3, 6, and 8. The capstone of the assessment program was the High School Graduation Qualifying Exam administered beginning in students' 10th-grade year. Although a comprehensive assessment program was now in place, it was only a matter of months before the requirements of NCLB (2002) led ADEED passing a new resolution asking that additional time be allowed to align the state's accountability as well as a provision for delaying the school designator program as efforts were made to bring the accountability system in alignment with the new requirements of NCLB (ADEED, 2007).

Although a new commissioner of education was appointed by Governor Murkowski in May of 2003, no significant changes were on the horizon until early in 2004, when the state announced that it had entered into a contract with Data Recognition Corporation (DRC) to develop a state-owned, integrated, and standards-based assessment (SBA) system at a savings in excess of $12 million over the existing system. In the following months, refinements were made to the state's assessment system, including regulations to clarify the use of accommodations and modifications on state assessments. These clarifications were the product of a class-action lawsuit, *Noon v. Alaska,* that resulted in a settlement in 2004 regarding the participation guidelines for students with disabilities. However, during this same period the results for Adequate Yearly Progress (AYP) as required by NCLB were released. Notably, only 42% of schools throughout the state met AYP targets. Although the second determination of AYP showed an increase to 58.8% of schools meeting AYP targets, the criteria for determining AYP had changed since the first determination; thus it was not clear to what extent the change affected the increase (ADEED, 2007).

In July of 2005, following the first administration of the new Alaska SBA, the state established standards for determining proficiency on the newly instituted Alaska SBA. In December, ADEED extended the Alaska SBA through the 10th grade. By July of 2006, proficiency levels were set for the 10th-grade SBA. At the same time, passing scores for the state's high school graduation qualifying exam were refined to more accurately reflect changes in standards and a general sense among educators that the score had previously been set too low (ADEED, 2007).

After several iterations and new initiatives, the final change to the progression of school reform came to fruition in 2006 when the U.S. Department of Education granted full approval to Alaska's assessment system as meeting the requirements of NCLB. According to ADEED (2007):

> Federal reviewers determined that Alaska's system aligned valid, reliable, and fair tests to challenging content and performance standards; involved a wide variety of Alaskans in developing the tests and standards; and used an effective, understandable method of presenting the results to the public. (p. 21)

The Alaska SBA is administered to students in Grades 3 through 10. The assessments are designed to measure the extent to which students have mastered the state's performance standards and the grade level equivalencies (GLEs) in reading, writing, and mathematics. The Alaska SBA subject tests were developed by DRC specifically for the state (ADEED, 2005). Student performance as measured by these assessments is used to determine an individual school's AYP as prescribed by NCLB.

With the provision in the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) that allows school districts to adopt Response to Intervention (RTI) as a means to identify students with learning disabilities, the use of Curriculum-based Measures (CBM) has been steadily increasing (Ardoin & Christ, 2008). Because of the large literature base for CBM and the efficiency with which the two tools can be used, both monetarily and time-wise, CBMs are an attractive alternative to teachers and administrators who understand the implications of using a nontraditional approach to assessing and predicting student achievement.

CBM Introduction

The origins of CBM can be traced back to the mid-1970s as an integral component of the Data-based Program Modification model (DBPM; Deno & Mirkin, 1977). The impetus for developing an efficient, valid, and reliable measure emerged from a desire to provide teachers with timely data for the purpose of evaluating the effects of their instruction. According to Fuchs and Deno (1991), perceived shortcomings noted during previous experiences with DBPM eventually led them to develop a general outcome model. For example, the DBPM model relied on mastery measurement of short-term instructional objectives. This reliance allowed teachers to have flexibility in adapting the model to their circumstances. In contrast, the general outcome measurement assesses general outcomes rather than specific skills (Fuchs & Deno, 1991). Equally important, general outcome measures rely on standardized practices that provide teachers with measures of student performance. Both mastery measurement and general outcome measures are examples of a broader category of assessment referred to as Curriculum-based Assessment (CBA; Fuchs & Deno, 1991).

The shift from assessing mastery of skill to measuring growth led to a decade or more of expanded use of CBA. Three ideas are central to the concept of CBA: (a) test stimuli are extracted from the local curricula, (b) students are tested multiple times, and (c) data from the assessments are used to inform instructional decision making (Fuchs & Deno, 1991). Although the concepts of CBA are straightforward, peculiarities of the various categories of CBA can be unclear; however, CBA generally falls into the two categories of mastery measurement and general outcome measurements, both of which are extracted from the local curriculum (Hintze, Christ, & Methe, 2006).

As researchers attempted to move assessment practices to curriculum based rather than skill based, CBM use emerged after 6 years of research (Deno, S., 1985). As researchers set out to develop CBM, four characteristics were identified as critical to the goal for teachers to be able to monitor student progress. S. Deno (1985) identified these four characteristics that were considered essential to the measures as the following: "[a] reliable and valid, . . . [b] simple and efficient, . . . [c] easily understood, . . . and [d] inexpensive" (p. 221). The movement away from assessment to measurement allowed for the development of standardized measures that were efficient, valid, and reliable; allowed for instructional decision making; and provided evaluative utility in assessing effectiveness of educational programs (Fuchs & Deno, 1991).

Although oral reading fluency is often overlooked, LaBerge and Samuels argued in 1974 that oral reading fluency was a good indicator of overall reading comprehension. Stanovich's (1984) interactive model, though it differs from the LaBerge and Samuels model, does hold to the assumption that oral reading fluency is an indicator of overall reading ability (Fuchs et al., 2001). Both models suggest that fluent low-level reading frees up higher level capacity that can be used for increased comprehension (Fuchs et al., 2001). This common assumption is the theoretical basis supporting oral reading fluency as an indicator of reading competence (Fuchs et al., 2001).

CBM Special Education Progress Monitoring

CBMs were initially developed for the purpose of measuring the effectiveness of special education programs by measuring student growth in specific academic areas

(Deno, S., 1985). Initially, CBMs were drawn directly from the curriculum. However, recent studies have shown that CBMs are valid indicators of student performance, even when the assessments are not drawn directly from the curriculum (Deno, S., 2003). This has ensured that CBM probes have been made commercially available from proprietary sources such as DIBELS by the University of Oregon and AIMSweb by Pearson.

Although CBMs were initially developed as mechanisms enabling teachers to monitor the effectiveness of their instruction (Deno, S., 2003), the use of CBMs has expanded considerably. Common or emerging uses of CBM include predicting performance against specific criteria, universal screeners for identifying at-risk students, developing local or site-based norms, evaluating the effect of pre-referral interventions, and identifying students with learning disabilities (Christ & Silberglitt, 2007; Deno, S., 2003). In addition to the value of using the CBM to identify at-risk students, CBMs can be given frequently, with reliability, and for regular progress monitoring of students targeted for tiered interventions or special education. CBM results provide teachers with regular feedback alerting them as to whether a specific intervention is working or needs to be changed. Multiple CBM measures could be given frequently to monitor the effectiveness of instruction and special education programs (Deno, S., 2003).

As previously noted in this chapter, CBMs do not need to be drawn directly from local curriculum (Fuchs & Deno, 1994). As a result, one commercially produced reading assessment used by school districts nationally is AIMSweb by Pearson. “AIMSweb® is a scientifically based, formative assessment system that ‘informs’ the teaching and learning process by providing continuous student performance data and reporting improvement to parents, teachers, and administrators to enable evidence-based evaluation and data-driven instruction” (AIMSweb, 2011a, p. 4).

AIMSweb (2011a) provides a Web-based database for subscribers to AIMSweb assessments. This database offers the ability to graph a student’s progress and project trend lines for expected student growth. In addition, the database houses assessment data from districts across the nation, allowing districts to compare their local students to national averages or to create their own district norms. This reporting allows districts to use CBM data for a variety of decision-making tasks (e.g., RTI and problem solving; Shinn & Shinn, 2002). Moreover, schools and districts have the ability “through web-based data management and reporting applications to provide a proactive and preventive solution for universal screening and progress monitoring for general education, strategic assessment for remedial programs or at risk, and intensive progress monitoring” (AIMSweb, 2011a, para. 4).

CBM High-stakes Relationship with Set Cut Scores and Norms

CBM might prove to be an alternative to traditional testing methods. Schools that have moved to a prevention-oriented system make use of universal screeners to track student progress as well as to assist in the identification of students who might be at risk of falling behind (Fuchs & Fuchs, 2006; Good et al., 2001). As noted by S. Deno (2003), the CBM is emerging as a possible alternative to the traditional assessment in that it might have predictive utility.

The review of literature revealed that several recent studies have begun to examine the relationship between student performance on R-CBM and reading performance on state high-stakes assessments (Crawford et al., 2001; Fuchs et al., 2001;

Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001; Wood, 2006). These studies identified reading fluency as measured by the R-CBM as showing significant correlations to state high-stakes assessments. The correlations were significant enough that that the R-CBM might have specific utility in predicting student outcomes on high-stakes assessments for reading.

Stage and Jacobsen (2001) completed a study in Washington examining the relationship between oral reading fluency and the Washington Assessment of Student Learning (WASL) for fourth graders. Stage and Jacobsen were among the first to provide evidence that the R-CBM might have the ability to predict student outcomes on a state assessment. In this study, students were administered the R-CBM in September, January, and May. Using a growth curve analysis, benchmark or cutoff scores were established using the fall, winter, and spring assessment scores of fourth-grade students from a single elementary school in the state of Washington (n = 173) and were compared with student performance on the WASL. The analysis revealed that R-CBM scores were able to predict WASL scores at a rate higher than the base rate of the sample being studied. The correlation between the student scores in the fall had strong correlations to student performance on the WASL in the spring.

Crawford et al. (2001) conducted a 2-year study with a cohort of students as they moved from second to third grade. Crawford et al. provided specific analysis of the relationship between the students' oral reading fluency and their scores on their state assessments in both reading and math. Two main research questions were addressed. The first looked at the strength of the relationship between a student's oral reading rate and that student's future performance on the state reading and math tests. Crawford et al.

further argued that because of the limited research demonstrating the use of CBM to predict outcomes on state assessments, should their study demonstrate a strong relationship between oral reading fluency and a state criterion-referenced test, additional credence would be given to the expansion of using CBM to predict student outcomes on pre-established benchmarks. The second question addressed what level of reading fluency in the second and third grades could predict students' third-grade reading and math scores on statewide assessments. The results revealed a moderate correlation between third grade oral reading rates and students' outcomes on statewide assessments (r = .60). Due to the small sample in this study, the researchers relied on norms established by Hasbrouck and Tindal (1992). Using these norms, 81% of students reading above the 50th percentile passed the statewide assessment.

In an effort to expand on the findings of previous research, McGlinchey and Hixson (2004) replicated Stage and Jacobsen's (2001) study with several variations. McGlinchey and Hixson conducted their study using the Michigan Educational Assessment Program (MEAP) reading assessment. The rationale for using this assessment was that the MEAP was part of the high-stakes assessment for the districts involved. Moreover, the study encompassed eight school years and a more diverse population. Based on previous research, 100 WCPM (words correct per minute) was used as a cut score. To further the diagnostic accuracy of the results, McGlinchey and Hixson employed the five statistical measurements of sensitivity, specificity, positive predictive power, negative predictive power, and overall correct classification. Of the students who read 100 WCPM or greater, 74% achieved satisfactory scores on the MEAP. The CBM scores derived from this study established the negative predictive power as 72%, which was an improvement to the base rate of 46%. Negative predictive power was the probability that a student was correctly classified as attaining a satisfactory score on the MAEP (McGlinchey & Hixson, 2004).

A growing number of studies have shown CBMs to be both a valid and reliable indicators of student skill level (Deno, Mirkin, & Chiang, 1982; Fuchs, Tindal, & Deno, 1984; Marston, 1989; Shinn, Good, Knutson, Tilly, & Collins, 1992). In recent years, several studies have been done specifically examining the value of CBM and their utility in predicting student success on state standardized high-stakes assessments (Fuchs et al., 2001; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Wood, 2006).

Crawford et al. (2001) examined the relationship between oral reading fluency and the Oregon statewide achievement tests. Crawford et al. included both second and third graders. Not only did they confirm a high correlation between oral reading fluency and success on state assessments, but they also found that 100% of students with oral reading fluency of 72 words or greater in the second grade passed the third grade Oregon reading assessment. Crawford et al. confirmed the findings from the 2001 Good et al. study. Good et al. used established benchmarks for oral reading fluency for the purpose of predicting student success on the Oregon state reading assessment. Crawford et al. further supported that oral reading fluency was a prerequisite skill for reading comprehension. Additionally, the predictive utility of using CBM to predict outcomes on high-stakes reading assessments was demonstrated. Third-grade students participating in the study who had oral reading fluency of 110 words or better on grade-level passages were likely to pass the Oregon reading assessment (Crawford et al., 2001). While

Crawford et al. (2001) found that all students achieving certain benchmarks in oral fluency also passed the state reading exam. Studies were found in the literature investigating the nature of a relationship between any R-CBM and the Alaska SBA.

CBM High-stakes Establish Cut Scores in Study

McGlinchey and Hixson (2004) replicated the 2001 Stage and Jacobsen study using CBM to predict success on the MEAP. McGlinchey and Hixson had similar findings, but the correlations were found to be even higher than in the previous findings by Stage and Jacobsen. McGlinchey and Hixson attributed the practice of high-stakes assessment often commencing in the third grade to the requirement of NCLB that all children should be fluent readers by third grade. In response to the mounting pressure for schools to prevent reading difficulties through early identification and as discussed in Chapter 1, affordable, technically adequate assessments need to be available that can efficiently and accurately identify struggling readers prior to the third grade. CBMs may be able to meet this challenge. Characteristics of CBMs, according to S. Deno (2003), are that they are technically adequate, time efficient, and easy to teach. These studies all showed the CBM to be an emerging and promising predictor of student achievement on high-stakes assessments (Fuchs 2004).

Districts have tried to make meaning of how to use CBM to monitor student progress and determine if students are academically on course (Fuchs & Fuchs, 2004). To date, most districts have relied on normative CBM cut scores and percentile scores in an effort to predict student performance. Historically, the R-CBM cut score defines the number of words read correctly in one minute that students need to attain in order to be considered proficient in reading. The use of cut scores can be traced back to S. Deno

(1985) who established such scores using local samples. Others, like Hasbrouck and Tindal (1992), argued that norms established from larger, more diverse populations could be considered more stable. Hasbrouck and Tindal established a set of norms over a 9-year period. Moreover, they asserted that other districts attempting to define norms from a local population would likely find similar results and therefore could save the time and expense by using their newly established norms. According to Silberglitt and Hintze (2005), these norms have been often quoted and used for instructional decision making.

By using cut scores or norms established in previous studies against which their students can be compared, compounded with the variability in defining proficiency, many districts still do not have adequate assessment tools for early identification of students who may not be on track. For this reason, districts that have resorted to using previously established cut scores normed against differing populations and criteria are now attempting to set cut scores based on their specific populations. This effort has been aided by the requirements of NCLB. The NCLB now provides states with a specific benchmark of proficiency to which cut scores can be set (Silberglitt & Hintze, 2005).

Although many school districts have made efforts to improve early identification of students likely to be at risk of failing state exams, uncertainty as to which cut scores to use when using CBM as a screening tool continues among educators. Regardless of how a cut score is determined, the goal is to determine where to draw the line on the vertical axis. This cut score placement classifies students' scores as predicted to pass or to fail the state exam. Silberglitt and Hintze (2005) addressed this problem by reviewing the methods of setting cut scores in previous studies in an attempt to determine the most advantageous method. Among the earliest studies reviewed by Silberglitt and Hintze was the 2001 study conducted by Good et al. (2001).

In this study, Good et al. (2001) maintained that not only should a benchmark be measurable, but it should also be tied to indicators of student performance. Presented as equally important, Good et al. argued that "a benchmark goal should be linked to or anchored by a socially meaningful and important outcome. Ideally, establishment of a benchmark goal integrates statistical psychometric and sociopolitical consideration in an overall judgment" (p. 266). They built on the work of the norms established previously by Hasbrouck and Tindal (1992). Moreover, Good et al. addressed two problems associated with using the norms established by Hasbrouck and Tindal. Good et al. recognized that norms from previous studies might be representative of performance within the representative sample but questioned whether the norms represented a rigorous enough threshold as an expectation of student performance. Further, Good et al. recognized the possible dynamic nature of using normative cut scores for goal-setting purposes:

> After all, the intent of a goal is to provide a target for all children to attain. However, if we have a normative-based target, and we are effective in reaching the target, the target will necessarily move. No matter how effective our instruction, 50% of children will still be below the middle performance. (p. 270)

In an effort to move from norms to cut scores, Good et al. (2001) anchored their study with a first-grade spring cut score of 40 WCPM. Justification for this score was detailed with the assertion that the score met their previously established criteria as having support from empirical, theoretical, and social-validation sources. Using 40 WCPM in spring of first grade as a starting point, Good et al. analyzed relationships to establish cut scores linking one benchmarking period to the next, culminating with the third-grade reading assessment.

A growing number of studies have demonstrated the relationship between CBM and state achievement tests (Crawford et al., 2001; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Sibley et al., 2007; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). In a recent meta-analysis of 14 state assessments, Yeo (2010) identified 27 studies demonstrating the use of CBM to predict outcomes on state achievement tests. Although Yeo revealed an overall correlation coefficient of .689, the range of correlations varied greatly between states. Yeo presented a possible explanation for the variations in that each state is required to create its own assessment, and "it is reasonable to expect the correlation between CBM and statewide achievement tests, as presented by research studies, may be heterogeneous between states" (p. 413). Due to the variability of correlation coefficients and the assertion that correlations were likely to be heterogeneous between states, Yeo recommended further research continue the investigation of the relationship between CBM and other state tests.

The ability to set cut scores or benchmarks with CBMs that can predict outcomes on state assessments has some distinct advantages. Establishment of benchmarks for each grade affords districts a tool for ongoing progress monitoring. Such monitoring can be used to alter instruction in preparing students for state assessments (McGlinchey & Hixson, 2004). As this body of research emerges, multiple statistical methods have been used to establish cut scores.

Good et al. (2001) used scatter plots to establish linkages and ultimately established cut scores that linked likelihood of attaining benchmarks from one grade to the next. McGlinchey and Hixson (2004) evaluated the relationship between CBM and student performance on the MEAP and used previously established cut scores in the study. Diagnostic efficiency statistics were then used to verify their accuracy in predicting student outcomes on the state assessment. Stage and Jacobsen (2001) used analysis of variance (ANOVA) and the hierarchical linear model (HLM) along with diagnostic efficiency statistics to establish cut scores and check them for diagnostic accuracy for predicting failure on the WASL.

Silberglitt and Hintze (2005) discussed the benefit of a consistent set of cut scores when determining whether students are on track to pass state assessments. Moreover, they recognized that measuring student growth relative to consistent cut scores also provides districts with an opportunity to assess the effectiveness of interventions and instruction. Lastly, Silberglitt and Hintze recognized that specific cut scores meet requirements for assessments critical to RTI constructs. Although CBMs have been previously established for determining nonresponsiveness, previous studies had relied on normative data (Silberglitt & Hintze, 2005). Silberglitt and Hintze noted:

> Cut scores developed by linking R-CBM to state test performance present an alternative to normative data for establishing these criteria. . . . Using cut scores linked to statewide assessments provides a consistent set of rigorous criteria for judging student performance. (p. 322)

Silberglitt and Hintze (2005) evaluated four methods for determining criterion referenced cut scores: (a) discriminant analysis (DA), (b) equipercentile method, (c) logistic regression, and (d) Receiver Operating Characteristic (ROC). Logistic regression was found to yield the highest level of diagnostic accuracy, but Silberglitt and Hintze ultimately used DA to identify a range of acceptable scores and used ROC analysis to maximize the number of false negatives through the use of a priori rules. By emphasizing sensitivity over specificity, they sought to ensure that students who were at or above the cut score would also pass the state assessment.

Shapiro et al. (2006) conducted a similar study using a combination of LA and ROC. They reported their findings as being consistent with previous research and demonstrated a relationship between CBM and the Pennsylvania System of School Assessment. In contrast to Yeo (2010), Shapiro et al. believed that even though each state was responsible for creating its own assessment, CBMs can predict outcomes on state assessments without regard to differences between assessments by state. Shapiro et al. stated:

> Considering that each state assessment measure is typically built to evaluate student progress toward competency on state curriculum standards and that these standards vary considerably from state to state, CBM is indeed a very powerful measurement tool that appears to transcend the differences in state assessments. (p. 28)

Response to Intervention

In Chapter 1, discussion regarding specific instruments having met specific criteria established for educational and psychological testing (e.g., is CBM valid and reliable?). AIMSweb (2011a) reported its R-CBM assessment met all seven criteria outlined for educational and psychological testing:

> Progress monitoring tools were evaluated according to the degree to which they met seven criteria derived from the Standards for Educational and Psychological Testing developed by the Joint Committee appointed by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement Used in Education (NCMUE) and the Individuals with Disabilities Education Act (IDEA). (para. 4)

RTI theory and main ideas. The term RTI is often confusing because of the different meanings and interpretations adopted throughout the country. IDEIA (2004) specifically references the allowance of processes that measure a student's response to interventions: "In determining whether a child has a specific learning disability, a local educational agency may use a process that determines if the child responds to scientific, research-based intervention as a part of the evaluation procedures" (sec. 614.b.6.B). Likewise, both the Office of Special Education (OSEP; Bradley & Danielson, 2004) and the National Association of State Directors of Special Education (NASDSE, 2008a, 2008b) referred to RTI. Several state departments of education, including Alaska's, use the term to reflect that intervention and instruction are interchangeable within the construct. Although no one universally accepted definition of RTI has achieved consensus (Kavale & Spaulding, 2008; National Joint Committee on Learning Disabilities [NJCLD], 2005), the principles can be cataloged as a process that uses interval data to measure a student's responsiveness to changes in instruction. Students who do not respond as expected are provided more intense interventions (Brown-Chidsey & Steege, 2010; Hunley, McNamara, & National Association of School Psychologists, 2010; Johnson, Mellard, Fuchs, & McKnight, 2006; NASDSE, 2005, 2008b; National Center on Response to Intervention [NCRTI], 2010). Given the lack of a common definition of RTI, the NCRTI (2010) provides a definition that captures the basic tenets of RTI:

> Response to intervention integrates assessment and intervention within a multi-level prevention system to maximize student achievement. . . . With RTI, schools use data to identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions, and adjust the intensity and nature of those interventions depending on a student's responsiveness. (p. 2)

An alternative definition was presented by Batsche et al. (2005), in which RTI was represented as the practice of "(1) providing high-quality instruction/intervention matched to student needs and (2) using learning rate over time and level of performance to (3) make important educational decisions" (p. 1). Without a mutually agreed-upon definition and structure for RTI, main tenets of RTI as an essential universally accepted core are difficult to identify.

RTI is largely viewed as a model built on a convergence of multiple initiatives that have emerged of the past two decades (Barnett, Daly, Jones & Lentz, 2004; NJCLD, 2005). Prior to the reauthorization of IDEA, major emphasis was placed on individualized instruction. IDEIA (2004) now allows the use of RTI in the determination of learning disabilities. Paired with NCLB embracing the principals of RTI, an emphasis within RTI models to address the needs of groups rather than individual students has emerged (Kavale & Spaulding, 2008). The reading research of the 1990s demonstrated that early intervention and preventative measures made significant differences to struggling readers. Intervention studies by the National Institute of Child Health and Human Development (NICHD) were in part responsible for an RTI approach being written in the IDEA 2004 (NJCLD, 2005). For this reason, through universal screening of all students, RTI focuses on at-risk students rather than those with deficits (Vaughn & Fuchs 2003).

Batche et al. (2005) identified "Deno's data-based program modification model (Deno, S., 1985; Deno & Mirkin, 1977) and Bergan's behavioral consultation model (Bergan, 1977; Bergan's & Kratochwill, 1990)" as major contributors to modern RTI practices (p. 7). Because RTI models embody many principals, the theoretical framework is not easily understood. Gallagher (2010) attempted to quantify the theory of RTI:

> The apparent logic behind the new *Response to Intervention* (RTI) model is reminiscent of the original learning disability theory in that it relies on inferring

> the presence of learning disability based on interpretation of behavior. Those who do not respond to instructional interventions "scientifically" proven to be effective must accordingly have the disability. (Response to Intervention: Learning Disability for the Twenty-First Century, para. 3)

Both Bergan's and modern day RTI problem solving models are extensions of behavioral and social learning theory (Castillo & Batsche 2012). The deficit model of Learning Disabilities relied on an IQ or ability component consistent with cognitive learning theory, RTI model rely on universal screening and progress monitoring data to determine the degree to which students need additional supports (Glover & DiPerna, 2007). This practice is consistent with behavioral learning theory (Ertmer & Newby, 1993). Though many aspects of RTI are grounded in behavioral theory, RTI appears to be an instructional model attempting to apply theory to practice.

History of RTI. December 3, 2004, President George W. Bush signed IDEA into law. This act has proven to be a catalyst in propelling the RTI forward in education. Although IDEA was instrumental by creating momentum in the RTI movement, the concept of RTI is not new. Many of the concepts of RTI have been researched and practiced dating back to the early 1970s. E. Deno (1970) wrote about the need to reform special education and presented a model of cascading services. This model was in stark contrast to the current practice and policies. Current practice favored a model presuming lack of progress by students to be rooted in an organic deficit. For this reason, both policies and practices currently relied on pathology when considering treatment. Moreover, E. Deno identified the need to move away from a deficit model and attributed student struggles to the very practices employed by educators intended to promote students with fulfillment of their own self-realization.

> [Our] viewpoint must switch from the present fix on pathology . . . to approaches which emphasize the fact that the problem is not in the child but in the mismatch

> which exists between the child's need and the opportunities we [special educators] make available. (Deno, E., 1970, p. 229)

Recognizing that there were limited measures to assess the optimization of a learning environment, E. Deno (1970) proposed a model she termed *cascade of services*. The model provided a mechanism so that rather than relying solely on diagnostic measures to determine eligibility, the practitioner could be involved determining movement between general and special education. By making services available to students based on need and allowing for some movement between levels in the cascade of services, it was believed that there could be a reduction in students formally identified as needing special education while at the same time students requiring additional services would still have access.

During the late 1970s, two bodies of research resulted in instructional models that incorporated the concept of dynamic allocation of services-based student needs. Both models, Deno's data-based program modification model (Deno, S., 1985; Deno & Mirkin, 1977) and Bergan's behavioral consultation model (Bergan, 1977) incorporated practices seen in RTI practices today. That is to illustrate that both models were based on similar principals that set RTI apart from other instructional models. In particular, both models required systemic implementation and the use of data in decision making. These models were based in part on the ideas of cascades of service outlined in the previous work of Deno (Brown-Chidsey & Steege, 2010).

According to Bender and Shores (2007), the use of an RTI approach in determining whether a student has a learning disability can be traced to a study conducted by the National Research Council. Heller, Holtzman, and Messick (1982) had deemed that the validity of processes used in the identification of categorical disabilities should be measured against three criteria. The first was whether one could expect students to make progress based on the quality of the core instruction. Second, special education programs needed to improve student performance in order to be warranted. Lastly, the evaluation and assessments to make classification decisions needed to me meaningful and accurate (Bender & Shores, 2007; Vaughn & Fuchs, 2003).

Moreover, Bender and Shores (2007) identified characteristics of RTI addressed by both Deno and Mirkin (1977) and Bergan (1977). Models in both of these studies held the following ideas in common: (a) the academic or behavior problem was identified and clearly written goals commensurate with student performance were developed in order address the concern, (b) an intervention plan was established that relied on evidence or research-based practices, (c) a curriculum-based assessment was used to monitor student progress, and 4 (d) determination of intervention effectiveness was data-driven relying on whether students met their goals (Bender & Shores, 2007). A side by side comparison of the salient characteristics is presented in Figure 2.

Both the DBPM model (Deno & Mirkin, 1977) and Bergan's (1977) problem-solving model evolved to resemble what is often seen in practice today. DBPM focused on specific academic skills, which were assessed using CBM. These measures were typically extracted from the local curriculum as a form of formative assessment.

Bergan Model and Modern Problem-solving Steps	Deno Model and Modern Standard Protocol Reading Interventions
Define the problem behaviorally.	Define the problems in terms of performance level and skills deficits.
Measure performance in the natural setting.	Access reading skills through progress-monitoring, CBM, and criterion-referenced skills inventories.
Determine current status and performance gap compared to peers.	Determine current status and performance gap compared to peers.
State a goal based on peer performance expectations.	State goals in terms of benchmarks for reading performance and peer expectations.
Design intervention plan, applying scientific instructional and behavior change principals.	Apply scientifically based instruction emphasizing five components of reading.
Implement intervention over a reasonable period of time with good treatment integrity.	Implement intervention over a reasonable period of time with good treatment integrity.
Monitor progress frequently using a time series analysis graph and make changes in the intervention as needed to improve effectiveness or raise goals, as indicated by data.	Monitor progress frequently using a time series analysis graph and make changes in the intervention as needed to improve effectiveness or raise goals, as indicated by data.
Evaluate results compared to goals and peer performance.	Evaluate results compared to goals and peer performance.
Make decisions based on data to continue, fade, discontinue, or seek more intense interventions.	Make decisions based on data to continue, fade, discontinue, or seek more intense interventions.

Figure 2. The Bergan and Deno models adapted from NASDSE (2005, p. 8).

The measures were sensitive to small changes in growth, allowing teachers to adapt instruction accordingly. Decision rules were established to assist educators in determining the effectiveness of specific interventions and whether or not students were on track to meet their goals. On the other hand, the Bergan model, known as a problem-solving model, relied on hypothesis testing through a systematic approach to address both academic and behavioral deficits (NASDSE, 2005).

RTI core principles. As previously addressed in this dissertation, there is no specific definition that has been universally agreed upon for RTI. Likewise, the essential components of RTI are not unanimously agreed upon. NASDSE and Council of

Administrators of Special Education (NASDSE &CASE, 2006) identified the following essential components and beliefs as integral to an RTI model:

1. Believe that we can effectively teach all children.
2. Intervene early.
3. Use a multi-tiered model of service delivery.
4. Use problem solving to make decisions within a multi-tiered model.
5. Use research-based, scientifically validated interventions/instruction to the extent available.
6. Monitor student progress to inform instruction.
7. Use data to make decisions. A data-based decision regarding student response to intervention is central to RTI practices.
8. Use assessment for three different purposes (screening, diagnostics, progress monitoring). (pp. 20-21)

Multi-tiered service delivery. Although the number of tiers within an RTI model is not definitive, most models found today consist of three (Bender, 2009; Berkley, Bender, Peaster, & Saunders, 2009; NJCLD, 2005; Vaughn & Fuchs, 2003; see Figure 3).

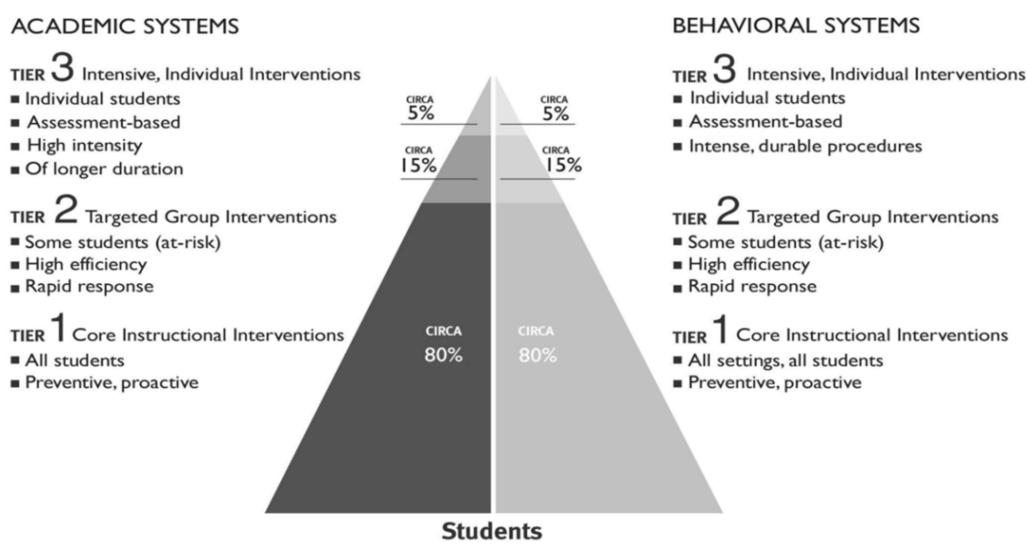


Figure 3. Three-Tier model of school supports from NASDSE (2005, p. 22).

Within an RTI construct, it is expected that the core curriculum should meet the needs of approximately 80% of students. Therefore, the base of the RTI triangle is representative of Tier 1 in that it must accommodate the largest number of students. Tier 1 consists of primary services or core curriculum. Within Tier 1, students receive quality instruction through a researched-based curriculum aligned to district or state standards. At Tier 1, the general education teacher is considered to be the interventionist and uses differentiated instruction, flexible groupings, and general accommodations to meet individual student needs. Another component of Tier 1 is universal screening. This allows districts to make quantitative judgments regarding the effectiveness of the core curriculum and to determine if Tier 1 is an appropriate placement. More importantly, universal screening identifies students likely to be in need of additional services. After verifying the efficacy and fidelity of interventions at Tier 1, students who are not succeeding can be referred to Tier 2. Secondary services are generally provided to students in addition to their primary services. Tier 2 services can be provided in the regular classroom or as supplemental services in another setting. Tier 3 or tertiary services are the most intensive services and are provided to students who have not responded to Tier 2 services. These services can occur in addition to Tier 1 services, but in extreme cases, they can also supplant the core curriculum.

Progress monitoring. Once students are placed in an appropriate intervention, they are regularly monitored at frequent intervals for the purpose of determining whether the intervention is effective. In this manner, teachers are provided with timely feedback and data that aid in determining the effectiveness of the intervention. Essential to RTI models is that districts have appropriate tools that are both valid and reliable when used in making placement and intervention decisions. CBMs have emerged as a tool that has met these criteria (Fuchs & Fuchs, 2004; Shinn, 2007). Accordingly, the CBM is commonly used as a progress monitoring tool within an RTI construct (Stecker, Lembke, & Foegen, 2008). Moreover, CBMs have proven to be effective for monitoring student progress both in the general education and special education environments (Deno, S., 2003; Shinn, 2007).

Data-based decisions. Most RTI models rely on established norms or establish local norms to assist in not only identifying students who are at risk but, equally important, establishing decision rules used in determining the intensity of intervention (i.e., Level 1, Level 2, or Level 3). With a focus on prevention, RTI models universally screen all students in an effort to identify students who are not on track with grade level expectations (Batsche et al., 2005; Shinn, 2007). Schools often rely on both normative and criterion data during the progress monitoring phase. Normative data established in large norming groups as well as cut scores established by individual states can be used to determine when a student is learning at rates commensurate with their same grade or age peers (Johnson, Jenkins, Petscher, & Catts, 2009; Silberglitt & Hintze, 2005). Even more important, using CBM to establish expected rates of growth over time provides teachers with the timely feedback needed to determine if a change to either instruction or the intervention is needed (Crawford et al., 2001; Deno, S., 1985). For this reason, progress-monitoring data are fundamental in determining whether a student's progress is sufficient to keep him or her on track and eventually catch up with his or her same age/grade peers.

RTI Models

Largely as a result of two bodies of research, two distinct models have emerged within an RTI construct: the problem solving and the standard protocol. Fuchs et al. (2003) identified two distinct camps of RTI proponents, each endorsing models that were developed largely through the efforts of practitioners in their respective fields: "an early intervention/prevention group consisting of early reading researchers and behaviorallyoriented school psychologists" (p. 159). While the school psychologists saw RTI as being synonymous with a problem-solving approach, the researchers strongly aligned to the standard protocol model (Christ et al., 2005; Fuchs et al., 2003). Because both have strengths and weaknesses, districts often use a combined approach when establishing their own models (Fuchs & Fuchs, 2006).

Standard protocol. The standard protocol model emerged from Deno's (1985) early work with data-based program modification (see Brown-Chidsey & Steege, 2010). The standard protocol model is prescriptive in that specific interventions are identified in advance to be used as interventions (McCook, 2006). Interventions are selected based on research-based practices supporting effectiveness. Standard protocol models focus largely on academic deficits, and it is customary practice to place students into small intervention groups based on common needs (i.e., reading comprehension; Bender & Shores, 2007; Fuchs & Fuchs, 2006). CBMs are used regularly to monitor students' performance throughout the intervention. Within this construct, judgments are not made on student performance in comparison to their peers but rather to their own growth over time compared to their prior performance (Bender & Shores, 2007; Fuchs & Fuchs, 2006).

RTI problem-solving process. In contrast, the problem-solving model can be characterized as inductive, empirical, and behavioral (Fuchs et al., 2003). Where the standard protocol model relies on predetermined interventions and decision rules, the problem-solving model caters specifically to individual students' needs. Adhering to a fundamental belief that neither categorical disabilities nor other student characteristics can determine appropriate interventions, proponents of the problem-solving model rely on the following four-stage problem-solving process: (a) defining the problem, (b) problem analysis, (c) developing and implementing a plan, and (d) evaluating whether or not the plan or intervention worked (Bender & Shores, 2007; NASDSE, 2005).

RTI hybrid model. Both the standard protocol model and the problem-solving model of RTI have inherent strengths and weaknesses. For this reason, most RTI models being adopted combine aspects of each. Hence, most RTI models are considered hybrid. (Fuchs, Mock, Morgan, & Young, 2003; McCook, 2006). In a hybrid model, evidence-based interventions are identified across the tiers. Based on universal screening data, intervention teams are able to determine if the problem is curriculum based, instructional, or student specific. If warranted, intervention teams then select an appropriate intervention based on student need. During the course of the intervention, a student's progress is monitored with CBMs so that a determination can be made regarding that student's responsiveness. Based on established criteria for grade and age performance, rate of growth, and length and intensity of individual interventions, teams review student data to determine whether the intervention has been effective (Burns & Gibbons, 2008).

Traditional methods of intervention are based on a presumption that the lack of performance is due to deficits in the child. In a hybrid model, first educators must establish that the curriculum is effective for the majority of students. Most models require that the core curriculum, if implemented with fidelity, meets the needs of approximately 80% of students. If a student progresses through each of the tiers of intervention while receiving appropriate instruction and interventions that are matched appropriately to a student's perceived need, only then does the focus shift to the individual child (McCook, 2006).

RTI Application to General and Special Education

Presently, the only law to make specific mention of RTI is IDEIA (2004). This has resulted in a widespread perception that RTI is a special education initiative (McMaster & Espin, 2007; Shores & Chester, 2009). According to the President's Commission on Excellence in Special Education (PCESE, 2002):

> The current system uses an antiquated model that waits for a child to fail, instead of a model based on prevention and intervention. Too little emphasis is put on prevention, early and accurate identification of learning and behavior problems and aggressive intervention using research-based approaches. This means students with disabilities do not get help early when that help can be most effective. Special education should be for those who do not respond to strong and appropriate instruction and methods provided in general education. (p. 7)

The resulting recommendations from PCESE played in integral role in new language being incorporated into IDEA and the reauthorized act through IDEIA. The IDEIA has provisions that allow LEAs to use the data collected through the RTI process in making eligibility determination for students suspected of having learning disabilities. While NCLB does not specifically mention RTI in its language, it does embody the principles of RTI throughout its language. Consequently, RTI has emerged over a period of four decades as part of improving upon special education.

The language incorporated into IDEIA makes it clear that RTI is a general education initiative. The importance of RTI was stressed in a joint paper by both NASDSE and CASE (2006), in which the use of RTI in general education settings was emphasized and the general education community was challenged "to join together to commit to a uniform system of education, where RTI plays a key role in identifying and working with struggling learners" (p. 2). This argument evolved from evidence that the identification of learning disabilities must originate in the general education classroom

(Brown-Chidsey & Steege, 2010; President's Commission on Excellence in Special Education [PCESE], 2002). Tilly (2003) argued that a three-tiered model of increasing intensity would provide the most effective architecture for this purpose. Most RTI models today consist of three tiers (Fuchs & Fuchs, 2007; NASDSE & CASE, 2006; Tilly, 2003).

Borrowing from the field of public health (Brown-Chidsey & Steege, 2010), most three-tiered models consist of a three-stage prevention model. Tier 1 consists of primary services or core curriculum. Tier 2, or secondary, services are for students who have not responded to primary services. Secondary services are generally provided to students in addition to their primary services. Tier 2 services can be provided in the regular classroom or as supplemental services in another setting. Tier 3, or tertiary, services are the most intensive services and are provided to students who have not responded to Tier 2 services. These services can occur in addition to Tier 1 services, but in extreme cases they can also supplant the core curriculum.

Local education agencies (LEAs), which embrace RTI as a means to improve education for all students, find that data collected through the RTI process can serve two important needs (Vanderheyden, 2011). First, schools that have established criterion referenced cut scores are able to evaluate overall effectiveness of their core instruction. Secondly, normative data can be used to ensure proper allocation of available resources; hence, general education shares the responsibility of RTI implementation and oversight (Burns & Gibbons, 2008; Flanagan, Ortiz, Alfonso, & Dynda, 2006; VanDerHeyden, Witt, & Barnett, 2005).

A measure of a student's success within an RTI construct is made possible by comparing aggregate grade-level student performance against expected performance (Johnson et al., 2009). Most commonly, districts have either adopted established norms (Hasbrook & Tindel, 1992) for this purpose or have developed cut scores (Silberglitt & Hintze, 2005) based on local criteria (i.e., scores that are reliable in predicting outcomes on high-stakes assessments). This process allows districts a mechanism for determining allocation of available resources. For example, if 80% of students are not meeting performance expectations, additional consideration must be given to the core curriculum and instruction. Consequently, the use of a three-tiered model provides an efficient mechanism by which educators can match resources to specific student needs (Fuchs & Fuchs, 2006; NASDSE, 2005).

RTI and Public Policy

IDEA. In 1977, IDEA became the first special education law. This law guaranteed free and appropriate education to all children with disabilities. The reauthorization of IDEA created a paradigm in which the emphasis of special education shifted from ensuring that students had access to services to an outcome-based model holding LEAs accountable for student performance (Bradley & Danielson, 2004; Individuals with Disabilities Education Improvement Act [IDEIA], 2004). Shortly prior to the enactment of the 1997 reauthorization of IDEA, the NJCLD expressed concerns to OSEP that further consideration needed to be given to the problems associated with the identification of children with learning disabilities. Further discussion was delayed until after the reauthorization of IDEA.

OSEP developed a four-step plan, later known as the learning disabilities initiative. The purpose of this initiative was to explore problems with existing methods of identifying learning disabilities. Following the 1997 reauthorization of IDEA, OSEP initiated a plan to include the formation of a diverse committee comprised of researchers, LEAs, policymakers, SEAs, parents, and trainers. This committee was to continue the conversation regarding existing concerns in the identification of learning disabilities (Bradley & Danielson, 2004).

The initial step was to commission nine papers exploring the concerns related to the identification of learning disabilities. Written responses were solicited for each of the nine papers. Specifically, one of the nine papers was dedicated to RTI. Second, a summit called Building a Foundation for the Future was conducted in August 2001 with the intent of accentuating the recent papers and furthering the discussion. Thereafter, a diverse group of stakeholders was amassed for the purpose of identifying future implications for research, policy, and practice. Finally, the researchers were asked to develop consensus statements for each of the papers to include a consensus statement for RTI. An abstract of this statement was presented by Bradley and Danielson (2004) as the following: "Response to scientifically valid and generally effective intervention is the most promising method of alternative identification and can both promote effective practices in schools and help to close the gap between identification and treatment" (p. 188). While IDEA did not mandate the use of RTI in determining the identification of students with learning disabilities, it did have a provision allowing states to use RTI in making such determinations.

It is often the case that the efforts of special education and general education are segregated and somewhat redundant. Often interventions are not matched to students' needs, in part due to lack of coordinated efforts between general and special education. Austin, Mattos, and Weber (2009) characterized the current relationship between general education and special education as symbolizing conflict and redundancies, lack of coordination, greater focus on paperwork and legal processes over results, and separate spheres of responsibility for teachers and students. In contrast, RTI provides a consistent, coordinated construct that allows educators to match interventions efficiently to specific student needs.

NCLB. A change in the paradigm of accountability in education took stage when President Bush signed NCLB into law in January of 2002. This latest reauthorization of the Elementary and Secondary Education Act (ESEA) placed great emphasis on state accountability systems and evidence-based practices. NCLB required that the progress of all students be measured against specific benchmarks. This requirement was to include students with disabilities, English language learning students, and students from low socioeconomic backgrounds (Burns & Gibbons, 2008; NCLB, 2001). The shift in accountability was inevitably tied to student outcomes on state assessments as a measure of student progress. NCLB further required states to provide evidence-based practices that could be validated through the use of data (Brown-Chidsey & Steege, 2010; Burns & Gibbons, 2008). In addition, NCLB required states to develop integrated assessment plans aligned to both content and achievement standards. Guidance from the U.S. Department of Education (2002) required the following:

> All states must submit plans to the Secretary of Education that include evidence that they have content and achievement standards and aligned assessments, school

> report card procedures, and statewide systems for holding schools and districts accountable for the achievement of their students. (p. 10)

Moreover, NCLB required state education agencies (SEAs) and LEAs to implement programs based on scientific research and effectiveness. Likewise, NCLB included a requirement that program effectiveness was to be monitored using valid and reliable data. The act further stipulated that the data should enhance instructional decisions leading to improved student performance (Brown-Chidsey & Steege, 2010; Burns & Gibbons, 2008; Shores & Chester, 2009).

The language in the NCLB emphasizing accountability for all students evolved from Deno's 1970 work regarding data-based decision making. The shift toward student learning outcomes and accountability for all students prescribed by NCLB "was endorsed by the Presidents Commission on Excellence in Special Education (2002) because 'those that get counted, count'" (Burns & Gibbons, 2008, p. 4).

School reform efforts and public policy during the previous decade have laid a foundation for RTI by incorporating many of the principles associated with RTI today. However, IDEIA was the first law to mention RTI specifically. For this reason, it is often believed that RTI's origins can be attributed to the signing of IDEIA. Specifically, IDEIA included a provision that allows states to use RTI as an alternative to the identification of students with learning disabilities. Although this provision has led to a perception that "RTI was born in special education law, it was conceived in the No Child Left Behind Act (NCLB) of 2001" (Burns & Gibbons, 2008, p. 4).

Although RTI evolved out of a special education initiative, provisions of RTI are more closely aligned to NCLB than to IDEA. This has left some in the IDEA camp wondering why special education policy is being influenced by general education. Yet, IDEA proponents have attributed positive results stemming from NCLB's increased emphasis on scientifically validated (evidence-based) interventions (Kavale & Spaulding, 2008).

Chapter 3: Methodology and Procedures

Overview

This chapter reiterates the purpose and research questions that serve as the framework for this study. Further, this chapter elaborates on the design, methodology, and subjects of the study. Within this section, specific discussion takes place regarding the subjects, data sources, and assurance to keep data and participants anonymous. A thorough description regarding the instrumentation along with an examination of the reliability and validity of instrumentation selected for each of the variables occurs. Last, a thorough discussion and detailed explanation of the data analysis and procedures used in conducting this study and subsequent summary concludes this chapter.

Purpose of the Study

The purpose of this study was to explore the relationship between student performance on R-CBMs and student performance on the Alaska's SBA administered to students in Studied School District Grade 3 through Grade 5 students in the Studied School District as required by Alaska's accountability system. Based on the statistically significant relationship between R-CBM and the Alaska SBA, the researcher examined the efficacy of deriving cut scores via logistic regression for use in predicting whether students were on track to meet proficiency requirements on the Alaska SBA. At the time of the study, Alaska was required to administer a criterion-referenced SBA and to all students Grades 3 through 10 as part of Alaska's educational accountability system. This study built on existing studies to verify previous findings further and to expand the current research by examining different populations.

Research Questions

Two research questions framed this study:

1. To what extent, if at all, is there a relationship between student performance on R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and student performance on the Alaska SBA administered in the spring of the same school year in the SSD?
2. To what extent, if at all, can cut scores be derived for each of the three R-CBM testing windows in the fall, winter, and spring that predict success on the Alaska SBA administered in the spring of the same school year in the SSD?

Research Design and Methodology

This non-experimental correlational study was divided into four phases. This approach was selected as a practical matter demonstrating that school districts might choose to establish criterion for identifying at-risk students using extant and readily available data. The first phase relied on descriptive statistics to assist in determining the normality and distribution of data. Second, the relationship between student performance on R-CBM tools and student performance on the Alaska SBA was examined using Pearson Correlation analysis. Third, binary logistic regression was used to determine cut scores for future use allowing educators to predict student outcomes on the Alaska Standard Based Assessment based on the same student's performance on their R-CBM. This phase of the study included the use of Receiver Operator Characteristic (ROC) curves to determine how well the logistic the model fit the data. Specifically the researcher investigated to what degree the model correctly classified success or failure of student performance on the Alaska SBA when using R-CBM as the predictor or independent variable.

This quantitative non-experimental study was retrospective in nature and relied on extant data to establish correlations between CBM and the Alaska SBA. R-CBM scores taken from three separate testing windows (fall, winter, and spring) during the school year were correlated to student performance on the Alaska SBA in the spring of the same year. Student performance on the Alaska SBA was artificially dichotomized into pass versus fail categories in order to achieve the purpose of the study. The use of cut scores for the predictor variable could be established for each testing period to predict success or failure accurately on the Alaska SBA ensured that the research questions could be answered.

Upon approval of the Pepperdine Institutional Review Board (IRB) and study school district (SSD), the researcher requested a census of all third, fourth, and fifth graders with any associated R-CBM scores for the 2007-2008, 2008-2009, 2009-2010, and 2010-2011 school years. Demographic information for each student was collected. The source for all data collection was limited to the SSD's student records system and/or the AIMSweb R-CBM database.

SSD and Participants

The SSD covered approximately 25,600 square miles. The SSD was among the top five largest school districts in Alaska and served approximately 9,500 students in pre-K through Grade 12. There were 44 separate schools in the district, of which 30% qualified for targeted Title I services. Approximately 14% of the students were eligible for special education services. Of the five largest school districts in the state, SSD was the only large district making Adequate Yearly Progress (AYP). The smallest school serving the elementary grades is a K-12 makeup with 10 students. The largest elementary school in the district served 450 students in pre-K through Grade 5.

During the time period of the data extraction for this study, SSD had 15 district office administrators, 45 site administrators, 610 certified staff, and 468 support staff. Of the 9,500 students enrolled in SSD, roughly 81% were classified as Caucasian, 10% as Alaska Native, 3% as Hispanic, 3% as multiethnic, and the remaining 4% as the sum of American Indian, Asian, Black, and Native Hawaiian/Pacific Islander. The 14 elementary schools varied in configurations, including Grades K-2, Grades K-6, Grades 3-5, and Grades 3-6 configurations. The SSD had four middle schools with either Grades 6-8 or Grades 7-8 configuration. There were 11 secondary schools consisting of Grades 7-12 and Grades 9-12 configurations. There were also 15 schools classified by SSD as *small schools* with the configurations of Grades 3-7, K-8, K-10, and K-12.

Archived R-CBM and Alaska SBA demographic and achievement data were collected for all Grades 3 through 5 students in the SSD who completed the SBA during one of the district's three universal screening assessments. Due to absenteeism and transiency within the population, it was expected that all students within the sample would not have R-CBM scores for all three testing periods. Data were evaluated to determine the appropriate method for addressing missing cases in the data set. Cases with missing values for one or more of the three R-CBM assessments were deleted list wise. The number of students included in the analysis was approximately 400 to 500 students per grade. The SSD was a semirural school district in south central Alaska. The ethnicity of the students in SSD remained relatively constant, and the demographics of the sample reflected past demographics.

Human Subjects Considerations

All students were assessed triennially with R-CBM as part of the SSD's existing assessment program. Likewise, all students were required to participate in the Alaska SBA in accordance with the state approved accountability plan. All assessments involved with this study would otherwise be conducted regardless of whether or not the study was occurring.

No data were recorded in such a manner that identifies or other demographic information would allow linking individual identities to their test scores. Data collected for this study were reported in aggregate form therefore no individual performance data or personal information were used or reported as a part of this study. Further, student data were only linked via a unique student ID provided by SSD. The identity of individual students was known only by SSD district officials. The researcher linked student Alaska SBA and R-CBM data via the district identifier but did not know individual students' identities. Only the researcher had access to the data collected for this study; therefore, informed consent was not required.

This study adhered to Pepperdine University's Institutional Review Board guidelines. Written permission to conduct the study was sought from SSD's superintendent according with SSD's guidelines prior to the commencing of any data collection for the study. All data and research materials collected for this study were kept confidential and secure with password protected files stored in locked file cabinets which were stored in a secure location. Three years after the conclusion of the study and publication of the study's results, all data will be destroyed. This study relied exclusively on extant data maintained by SSD, and this study had no direct impact on the subjects. Further, the researcher maintained anonymity of data to minimize any risk to human subjects.

This study was used to assist SSD in the evaluation of their instructional programs and student progress toward proficiency on the Alaska SBA. It was anticipated that in addition to assisting SSD evaluate their status towards making AYP, the results would be useful in evaluation of school level and individual student performance against a specific criteria. Further, this study contributed to a growing body of research focused on the utility of using R-CBM within an RTI construct.

Data Collection Procedures

Current data systems within the district provide unique codes to specific student records. However, the SSD issued random case numbers for the student cases used in the study to protect students' individual identities from the researcher. This working copy of the data was available to only the researcher and could not be tied directly or indirectly to individual students. The data obtained from SSD were stored on the researcher's computer and backed up to a secondary hard drive. Both of which are password protected and stored in a locked office and locked file cabinet for the duration of the study. Three years after the conclusion of the study and publication of the study's results, all data were destroyed. No names or unique identifiers were used in the study that would allow students to be linked to specific data. Large sample sizes provided added security to student identity. All data remained secure, and the researcher adhered to strict federal, state, and district guidelines.

This study used the following procedures to collect and analyze student data during the study:

1. The researcher met with the superintendent of SSD to review the scope of the proposed study, research questions, and methodology. Procedures for human subjects were reviewed to ensure that all data and student information are kept confidential. SSD procedures for securing permission to use SSD data were discussed. Following the initial meeting, a letter from the Superintendent granting permission to conduct the study was secured. This letter was sent to Pepperdine's IRB review committee along with the researchers IRB review application.
2. Subsequent to the IRB approval from Pepperdine, a data request was made through the SSD Superintendent. The scope of the request include all students in Grades 3 through 5 who completed the Alaska SBA and at least one R-CBM score that corresponded to the Alaska SBA testing year. Data were collected for the FY08 through FY11 school years. Specific information requested included Special education status, gender, LEP, school location, a unique student identifier, ethnicity, grade level testing year, and reading scale scores for the Alaska SBA and R-CBM scores for each the three screening windows.
3. The data were drawn from multiple sources, including SSD's student information system and the AIMSweb database. Prior to the data submission to the researcher, SSD aggregated the request into a single Excel file. SSD applied an algorithm to the student ID resulting in a unique student ID that was known only to the individual that prepared the data for the study. A single encrypted Excel file was emailed to the researcher with instructions to call for a password to unlock the file.

4. An initial review of the data indicated that SSD's data was not complete in the earlier years. It was further determined that the sample in later years was large enough to establish statistically derived cut scores. Consequently, analysis was completed using the fiscal year 2010 (FY10) data to derive cut scores and the fiscal year 2011 (FY11) data was used to complete the cross validation.
5. Prior to completing any analysis using SPSS version 17, the researcher completed a review of the Excel data file to determine if outliers in the data were present. Specifically score values outside of the scope of appropriate values (0-600) for the Alaska SBA. Cases with missing values for R-CBM were deleted list wise. SSD screens all students three times per year. It is expected that there will be missing scores due to transiency rates and attendance therefore missing cases were considered to be Missing Completely at Random. Four schools were deleted from the sample due to nonconforming testing calendars. Connections home school data were deleted from the sample because student instruction through this educational program was not a direct function of SSD.
6. A separate Excel file was created for the FY10 and FY11 school years. Within each workbook, a separate worksheet was created for the three R-CBM testing windows. Each worksheet included all demographic data, Alaska SBA reading scaled scores, and the R-CBM score for the specific administration.
7. Further data coding and analysis for both the FY10 and FY11 data were completed using SPSS version 17. The data provided in Excel files were imported to SPSS. School, gender, ethnicity, special education status, and limited English proficient were recoded into the same variable and given a numerical value as well as a

descriptive label should the new variable be needed for later reference. Scaled scores for the Alaska SBA were dichotomized and recoded into a separate variable (0 = fail, 1 = pass).

8. Descriptive statistics were obtained using the Frequency function in SPSS. Statistics were collected for Grades 3 through 5 and included the grade level reading Alaska SBA and the respective R-CBM for the three screeners administered in the same year.
9. A separate bivariate correlation analysis was completed between each grade level Alaska SBA reading scaled score and associated same grade R-CBM fluency scores obtained during the triennial screening process. This analysis included the Pearson coefficient and a one tailed test of significance. As previously discussed in this chapter, previous studies have consistently demonstrated positive correlations between R-CBM and state exams. Given an expectation that there would be a positive correlation, a one tailed test was used. A total of nine separate analyses were completed.
10. Binary logistic regression analysis was completed using dichotomize values of Alaska SBA (0 = not proficient, 1 = proficient) and R-CBM as the covariate. Predicted probabilities as well as predicted group membership were saved as part of this analysis. Options included in the analysis included Hosmer-Lemeshow goodness-of-fit, and CI for the odds ratios (i.e., exp(*B*)) set to 95%. A constant was included in the model.
11. A comparison of the means was conducted for each administration of R-CBM and the statistical likely hood that a specific R-CBM score was determined to predict that a student would pass the Alaska SBA. A predicted likelihood of 80% was used to

determine appropriate cut scores for each administration of R-CBM at each grade level.

12. Visual representation of the process and outcome was created by creating a scatter plot of predicted probabilities on the y axis and the respective R-CBM scores on the *x* axis. A horizontal reference line was added to the y axis at the .8 probability; a vertical reference line was placed at the corresponding R-CBM score identified to correspond to the average of means for predicted probabilities of .8. The cut score for each R-CBM administration is identified where these two lines intersect.
13. In order to assess how well the Logistic Regression model fit the data, ROC curves were generated for each R-CBM administration. R-CBM was plotted as the *Test Variable* and the dichotomized outcome variable was plotted as the *State Variable*. The value for the State Variable was set to 1. Options in the analysis included a diagonal reference line representing chance that was set to .50. The total area under an ROC curve (AUC) provided an indication of overall diagnostic accuracy. Values closer to 1.0 offered outstanding discrimination; values between .8 and .9 offered excellent discrimination; values between .7 and .8 offered acceptable discrimination, values closer to .6 offered questionable discrimination; but a value of .5 or less indicated that the predictor variable utility was no better than chance (Minitab, 2010).
14. Cross validation of cut scores determined through the binary logistic regression was completed with a multi-step process. R-CBM cut scores established via the comparison of means were re-coded in as dichotomized outcomes and save as a new variable. Values below 80% probability were coded as 0 for predicting non-

proficient on the Alaska SBA and values at or above 80% probability were coded as 1 for predicting proficient on the Alaska SBA.

15. A cross tab calculated using Alaska SBA dichotomized outcomes against the dichotomized values for the R-CBM.
16. Values were placed into an Excel spreadsheet and *True Positive, False Positive, True Negative,* and *False Negative* values were calculated. Based on these values, Specificity - Positive Predictive Power (PPP), and Sensitivity – Negative Predictive Power (NPP) and overall correct classification were calculated. True Positive refers to students who failed the ASBA and were predicted to fail. True Negative to students who passed the ASBA and were predicted to pass. False Positive refers to students who were predicted to fail the ASBA but passed. False negative refers to students who were predicted to pass the ASBA but failed. See Figure 4.

<table>
<tr><td colspan="2"></td><td colspan="2">R-CBM Test Status</td><td></td></tr>
<tr><td colspan="2"></td><td>Positive - Predicted Fail</td><td>Negative - Predicted Pass</td><td></td></tr>
<tr><td rowspan="2">Alaska SBA Status</td><td>Passed</td><td>False Positive - FP</td><td>True Negative - TN</td><td>Diagnostic Specificity
DSp = $\frac{TN}{TN + FP}$</td></tr>
<tr><td>Failed</td><td>True Positive - TP</td><td>False Negative - FN</td><td>Diagnostic Sensitivity
DSn = $\frac{TP}{TP + FN}$</td></tr>
<tr><td colspan="2"></td><td>Positive Predictive Power
PPP $\frac{TP}{TP + FP}$</td><td>Negative Predictive Power
NPP $\frac{TN}{FN + TN}$</td><td></td></tr>
</table>

Figure 4. Formulas for calculating diagnostic accuracy statistics.

17. The FY2011 values were compared to the values obtained in the binary logistic regression model established using the previous year's FY2010 data.

Instrumentation

R-CBM. The instrument used in this study was the AIMSweb® R-CBM, a test of student oral reading fluency. AIMSweb (2011a) is produced by Pearson Education for determining whether students should be included in RTI. All student performance data on the R-CBM are stored directly in the AIMSweb database.

Targets or benchmarks are established by subscribing agencies for the number of words students should be able to read correctly. The SSD currently uses the 25th percentile based on AIMSweb national aggregate norms as an indicator that students are at risk. This comparison is completed three times per year following the administration of R-CBM in each of the established testing periods. Hintze and Silberglitt (2005) established the validity of the AIMSweb R-CBM as a measure of oral reading fluency. Alternate form reliability was found to be .89 (Tindal, Marston, & Deno, 1983). Test-retest reliability was found to be .89 to .94 (Shinn, 2007). Moreover, the AIMSweb R-CBM was found to meet seven out of seven criteria for progress monitoring tools in a recent study funded by the U.S. Department of Education:

> Each progress monitoring tool that was submitted by publishers against these seven standards, (1) sufficient number of alternate forms with evidence of equal difficulty, (2) rates of improvement specified, (3) Benchmarks specified, (4) evidence of improved student learning or teacher planning, (5) sensitivity to student improvement, (6) reliability, and (7) validity, was judged independently by two of six members of the National Technical Review Panel. . . . Two AIMSweb Curriculum-Based Measures of Reading (R-CBM and Maze) fully met these seven standards. (AIMSweb, 2011a, para. 4-5)

During each R-CBM benchmarking or screening period, each student is asked to read aloud three separate reading passages. Sample text for the Grade 3 R-CBM is reproduced as Appendix A; of note, the R-CBM text for Grades 4 and 5 have not been available for reproduction. Students read for one minute from each passage, and the number of words read correctly is the oral reading fluency score. The median score for each student is recorded. The use of R-CBM for universal screening within Multi-Tiered Systems of Supports (MTSS) such as RTI allows students to be compared measured against existing criteria. Emerging in popularity is the ability to correlate R-CBM scores to criteria such as high-stakes assessments (Burns et al., 2002; Christ & Silberglitt, 2007). Standardized processes have been established for the use of R-CBM (Shinn, 2007). SSD follows standard procedures for R-CBM as part of its formative assessment and screening efforts.

Teachers responsible for the administration of the R-CBM were trained in the proper administration and scoring of the assessments. R-CBMs are administered triennially during three assessment windows, each 2 weeks in duration. The periods were fall (September 4-21), winter (January 22-February 1), and spring (April 28-May 9). Students were taken to a testing location on site and tested by a qualified assessor. Each assessment required less than five minutes to complete per administration.

Alaska standards-based assessment (SBA). The Alaska SBA is a high-stakes assessment designed to meet the requirements of NCLB. The assessment was developed for the state of Alaska by Data Recognition Corporation (ADEED, 2011b). Alaska's SBA is a criterion-referenced assessment administered to every student in Grades 3 through 10. The Alaska Grade Level Expectations (GLEs) are the criteria used for the Alaska SBA performance measurement (ADEED, 2011b). Most students will require approximately two to four hours to complete the reading assessment of the Alaska SBA (ADEED, 2011b). Reading is the only content area administered during a single school day.

The test was specifically designed to bring Alaska's assessment system into alignment with the requirements of NCLB. The assessment measures the degree to which students meet expectations on statewide performance standards. The reading portion of the Alaska SBA measures GLEs with a combination of Multiple-Choice (MC) and Constructed-Response (CR) Items. CR items are further delineated in to Short Constructed Response (SCR) and Extended Constructed response (ECR) items. For each MC item that a student answers correctly, they are awarded one point. For each SRC answered correctly, students are awarded from one to two points. ECR items are awarded from one to four points. The Alaska SBA Grade 3 practice reading assessment is provided in Appendix B as a complement to the Grade 3 R-CBM.

The assessment consists of 55 items: 52 multiple choice (MC) items, two 2-point constructed response (CR) items, and one 4-point CR item. Raw scores representing the number of correct MC responses plus total points from the CR items are converted to scale scores. Results of the SBA are reported in three categories: word identification skills, forming a general understanding, and analysis of general content or structure. A score of 300 has been established as proficient for Grades 3 through 10. Scale scores for all grades range from a minimum of 100 to a maximum of 600. Proficiency has been set by the state education agency to be a minimum of a score of 300 and is a standardized value used with each grade for the score needed for passing the Alaska SBA. The use of scale scores with a fixed measure of proficiency was established to allow year-to-year comparison of student performance relative to grade-level standards. Table 1 illustrates both the raw score and scaled scores for each for the proficiency levels.

Table 1

Reading Raw and Scale Score Cut Points for Each Proficiency Level

Level	Raw score cut point	Scale score cut point
Far Below Proficient	< 15	< 251
Below Proficient	15	251
Proficient	24	300
Advanced Proficient	48	418

All students in Grades 3 through 10 are expected to participate in the SBA. The only exception is for students with significant cognitive impairments, which participate in the alternate assessment. The SBA was administered by all school districts in Alaska over a 10- day window. During the 2010-2011 school year, the testing window ran from April 4, 2011 through April 18, 2011. Since the Alaska SBA testing materials are considered secure, training, administration, and security measure are in place and followed as outlined by DRC, ADEED, and the SSD. Once testing is completed, test administrators organize the assessments and send them to the district office where they are further packaged for shipping to the manufacturer of the test, Data Recognition Corporation (DRC), for scoring.

According to the state of Alaska and the test developer (ADEED, 2011b), the SBA has been deemed as content valid and reliable when used as a measure of student performance on GLEs. Students who test proficient on the assessment have been considered as performing at or near grade-level expectations (ADEED, 2005). The Alaska SBA was developed as an extension of the content to be assessed and is aligned directly to the Alaska content standards (ADEED, 2011b). Periodic review of the SBA is conducted by a committee of educators with the charge of ensuring content validity of the SBA. The committee continually assesses the extent to which the SBA is aligned to the Alaska content standards. If the committee determines that any item is not acceptable, they may elect to offer a revision or remove the item from the testing pool. "The nature and specificity of these review procedures provide strong evidence for the content validity of the SBAs" (ADEED, 2011b, p. 68).

Data Analysis Process

Being non-experimental in design, this study did not involve a control group. The assessments used in this study are required of schools and are an integral part of the district assessment program. Parents and students are notified of the timelines for the assessments through the school assessment matrix that is published and distributed each fall. The district testing matrix outlines each required assessment for students throughout the school year with the specific timelines and windows for each assessment. The matrix is distributed to all administrators in the fall and published on the district website. The matrix is also included in the student handbook, which is provided to each student. Parents sign that they have received the student handbook.

Properly trained staff members at respective sites were responsible for administering R-CBM each school year. Staff members at the respective schools were responsible for the proper scoring and entering of the student data from R-CBM into the AIMSweb database (Shinn & Shinn, 2002). Once data were entered into the database, they were available for the researcher for review. All students, whether or not they are participating in the study, are expected to participate in the assessments. The Alaska

SBA is administered to all students in Grades 3 through 10. Test coordinators and proctors are trained annually in accordance with the district, state and testing company, DRC expectations.

Organization and analysis of data was completed using a combination of Microsoft Excel, and SPSS software. The data analysis included a multistep process that determines whether R-CBM is a valid and appropriate tool for this purpose. Prior to examining the efficacy of establishing statistically derived cut scores, the correlation between R-CBM and the state SBA was examined. The Pearson *r* correlation coefficient was used to determine the relationship between student performance R-CBM tools and the Alaska SBA. Pearson *r* was used to determine the relationship between the two continuous variables. The degree of correlation assisted in determining whether R-CBM was a valid tool for use as a predictor for the dependent variable selected for this study. Descriptive statistics, Pearson correlation, and Logistic Regression were used in determining the validity of R-CBM as a predictor of student performance on the Alaska SBA.

Scatter plots provide a convenient way to evaluate R-CBM cut scores used in predicting outcomes on state exams. Scatter plots divide cut scores into four distinct quadrants (Figure 5). Each quadrant is defined by the intersection of a vertical line originating at the established R-CBM cut score on the horizontal axis and a horizontal line originating on the reading scale score on the Alaska SBA that demonstrates proficiency above the line and no proficiency below the line. In setting cut scores, it is desirable to minimize scores that result in excessive false negatives in that these students would be predicted to pass the Alaska SBA and therefore would not likely be receiving additional supports. Figure 5 illustrates that a R-CBM cut score appears to be a valid predictor of reading achievement on the Alaska SBA. Further, R-CBM scores yield high percentages of overall correct classifications (91.36% in this example with 467 true negatives and 51 true positives).

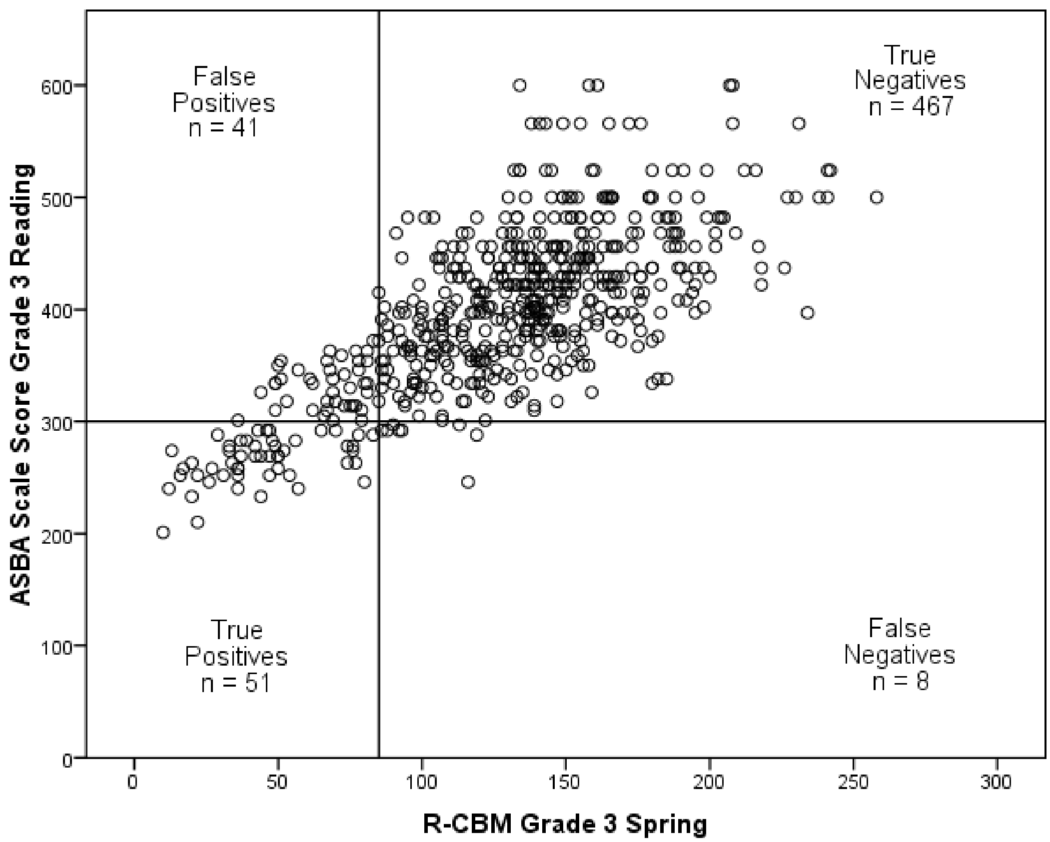


Figure 5. Scatter plot demonstrating R-CBM and Alaska SBA cut sores.

Figures 6 and 7 illustrate the use of diagnostic accuracy of R-CBM cut scores. Diagnostic accuracy is commonly evaluated using the following statistics: (a) *sensitivity* which refers to the students that failed the Alaska SBA that were predicted to do so by the R-CBM cut score; (b) *specificity* which refers to the students who passed the Alaska SBA as predicted by the R-CBM cut score; (c) *positive predictive power* which refers to the percentage of students predicted to fail the Alaska SBA that in fact do fail; (d) *negative predictive power* refers to the students who the student who pass the Alaska SBA which in turn pass (e) *True Positive* refers to students who failed the ASBA and were predicted to fail (f) *True Negative* to student who passed the ASBA and were predicted to pass; (g) *False Positive* refers to students who were predicted to fail the ASBA but passed; (h) *False negative* refers to students who were predicted to pass the ASBA but failed (McGlinchey & Hixson, 2004; Shapiro et al., 2006; Silberglitt et al., 2006; Stage & Jacobsen, 2001; Swets, Dawes, & Monahan, 2000).

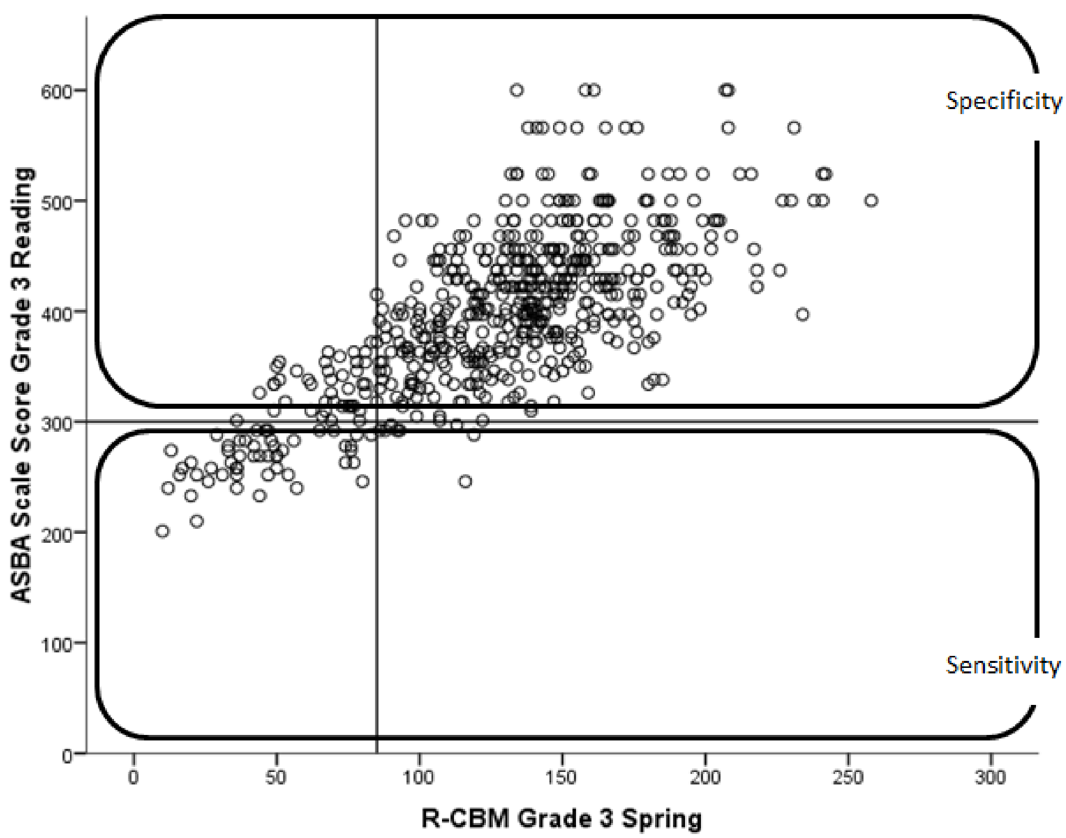


Figure 6. Scatter plot demonstrating sensitivity and specificity.

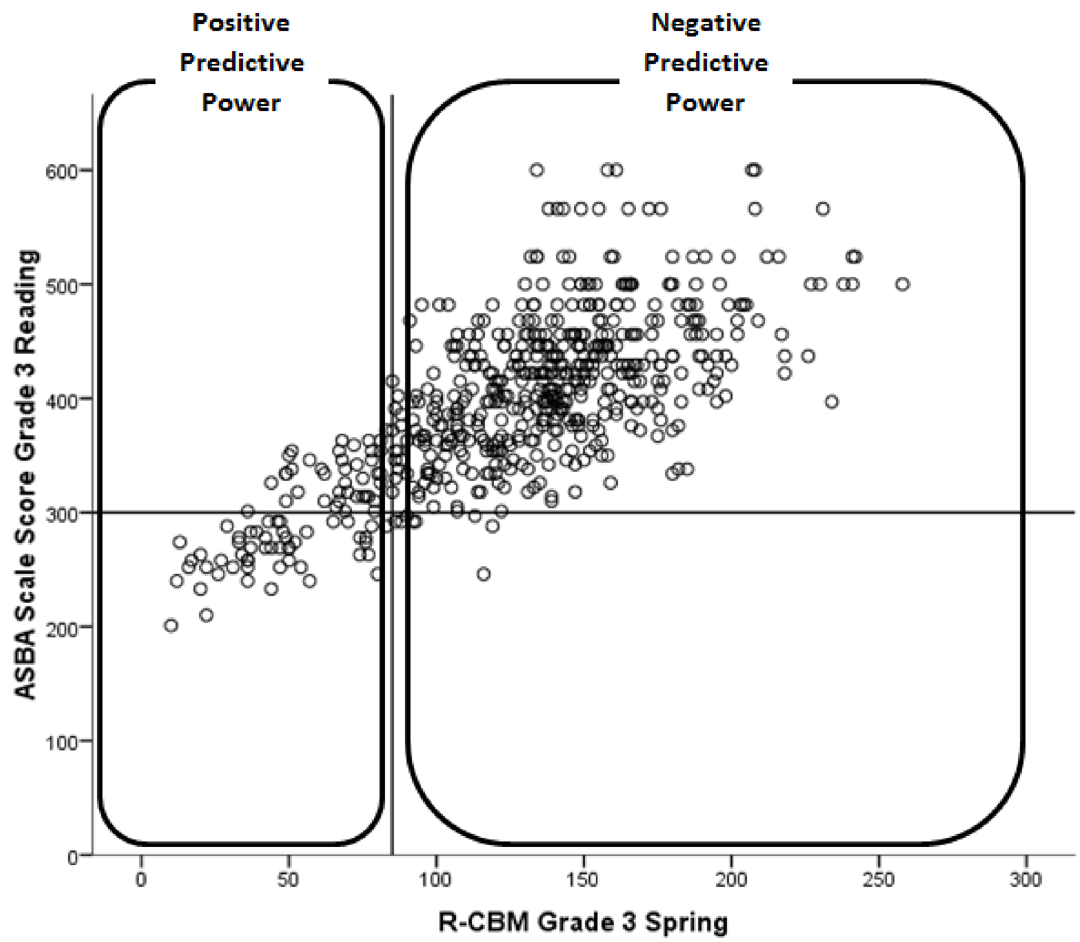


Figure 7. Scatter plot demonstrating positive and negative predictive power.

Chapter 4: Results

The purpose of this study was to explore the relationship between student performance on R-CBMs and student performance on the Alaska's SBA administered to students in Studied School District Grade 3 through Grade 5 students in the Studied School District as required by Alaska's accountability system. Two broad research questions framed this study across Grades 3, 4, and 5. Within each grade, the research questions were applied to interval scaled data obtained through the triennial administration of a curriculum-based measure of reading (R-CBM) in the 2009-2010 school year. The research questions were:

1. To what extent, if at all, is there a relationship between student performance on R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and the Alaska SBA administered in the spring of the same school year in the SSD?
2. To what extent, if at all, can cut scores be derived for each of the three R-CBM testing windows in the fall, winter, and spring that predict success on the Alaska SBA administered in the spring of the same school year in the SSD?

Descriptive Statistics

Descriptive statistics were calculated for the sample of third grade students (n = 472) who participated in the Alaska SBA during the 2009-2010 school year. The analysis included only students who completed the Alaska SBA and a fall, winter and spring R-CBM. The summary of student performance data is presented in Table 2. The normality of distribution was evaluated for each R-CBM and the Alaska SBA by examining the skewness and kurtosis of each distribution. (For histograms of distributions, see Appendix C.) All values were found to be between -0.58 and 0.20.

Standard distribution for R-CBM remained relatively constant at approximately 39 to 40 while the standard distribution for the Alaska SBA was 65.40 indicating a wider distribution of scores than found with the R-CBM. Means for R-CBM indicate that on average students reading increased by approximately 42 words between the fall and spring R-CBM screening.

Table 2

Descriptive Statistics of Third Grade Scores on R-CBM and the Alaska SBA in 2009-2010

3rd Grade Assessment	*n*	*M*	*SD*	Skewness	Skewness *SE*	Kurtosis	Kurtosis *SE*
Fall R-CBM	472	86.73	39.346	.199	.112	-.576	.224
Winter R-CBM	472	113.20	39.184	-.043	.112	-.223	.224
Spring R-CBM	472	128.28	40.190	-.099	.112	.045	.224
SBA Reading SS	472	394.23	65.394	-.249	.112	-.036	.224

Descriptive statistics were calculated for the sample of fourth grade students (n = 435) who participated in the Alaska SBA in 2009-2010. The analysis included only students who completed the Alaska SBA and a fall, winter and spring R-CBM. The summary of student performance data is presented in Table 3. The normality of distribution was evaluated for each R-CBM and the Alaska SBA by examining the skewness and kurtosis of each distribution. (For histograms of distributions, see Appendix C.) All values were found to be between -0.21 and 0.29. Standard distribution for R-CBM remained relatively constant at approximately 39 to 40 while the standard distribution for the Alaska SBA was 72.33 indicating a wider distribution of scores than found with the R-CBM. Means for R-CBM indicate that on average students reading increased by approximately 33 words between the fall and spring R-CBM screening.

Table 3

Descriptive Statistics Fourth Grade Scores on R-CBM and the Alaska SBA in 2009-2010

4th Grade Assessment	*n*	*M*	*SD*	Skewness	Skewness *SE*	Kurtosis	Kurtosis *SE*
Fall R-CBM	435	109.35	39.765	.149	.117	.040	.234
Winter R-CBM	435	130.29	40.434	.068	.117	.289	.234
Spring R-CBM	435	142.14	42.235	-.030	.117	.164	.234
SBA Reading SS	435	402.78	72.327	-.208	.117	.252	.234

Descriptive statistics were calculated for the sample of fifth grade students (*n* = 517) who participated in the Alaska SBA in 2009-2010. The analysis included only student who completed the Alaska SBA and a fall, winter, and spring R-CBM. The summary of student performance data is presented in Table 4. The normality of distribution was evaluated for each R-CBM and the Alaska SBA by examining the skewness and kurtosis of each distribution. All values were found to be between -0.10 and 0.31. (For histograms of distributions, see Appendix C.) Standard distribution for R-CBM remained relatively constant at approximately 41 to 42 while the standard distribution for the Alaska SBA was 63.33 indicating a wider distribution of scores than found with the R-CBM. Means for R-CBM indicate that on average students reading increased by approximately 32 words between the fall and spring R-CBM screening.

Table 4

Descriptive Statistics Fifth Grade Scores on R-CBM and the Alaska SBA in 2009-2010

5th Grade Assessment	*n*	*M*	*SD*	Skewness	Skewness *SE*	Kurtosis	Kurtosis *SE*
Fall R-CBM	517	128.16	41.967	.266	.107	.028	.214
Winter R-CBM	517	146.48	41.832	.087	.107	.251	.214
Spring R-CBM	517	160.96	42.310	-.090	.107	.234	.214
SBA Reading	517	397.24	63.332	-.102	.107	.314	.214

Correlations for Assessments Within Each Grade

The first research question sought to determine the extent of the relationship, if any, between student performance on R-CBM and student performance on the Alaska SBA administered in Grades 3, 4, and 5 in the fall, winter, and spring in the SSD. A Pearson product moment correlation was calculated for Grades 3 through 5 with R-CBM defined as the independent variable and the Alaska SBA defined as the dependent variable. A one-tailed test of significance was used for this analysis. Outcomes for each grade were reported separately.

Grade 3 results demonstrated that all correlation coefficients were significant at $p < .01$ for each administration of R-CBM (Table 5). There was a statistically significant positive correlation between the fall R-CBM administration ($M = 86.73$ $SD = 39.35$) and Alaska SBA ($M = 394.23$, $SD = 65.39$), $r = .689$, $p < .001$, $n = 472$. The winter administration resulted in slightly stronger correlations ($M = 113.20$, $SD = 39.18$) and Alaska SBA ($M = 394.23$, $SD = 65.39$) , $r = .700$, $p < .001$, $n = 472$. The strongest correlations were found in the spring administration (*M* = 128.28, *SD* = 40.19) and Alaska SBA (*M* = 394.23, *SD* = 65.39), $r = .728$, $p < .001$, $n = 472$.

Table 5

Correlations Between Third Grade Assessment Administrations

3rd Grade Assessment	SBA Reading SS	Fall R-CBM	Winter R-CBM	Spring R-CBM
SBA Reading SS	1	.689**	.700**	.728**

Note. $n = 472$. ** Correlation is significant at the 0.01 level (1-tailed).

Grade 4 results demonstrated that all correlation coefficients were significant at $p < .01$ for each administration of R-CBM (Table 6). There was a statistically significant positive correlation between the fall R-CBM administration (*M* = 109.35, *SD* = 39.77) and Alaska SBA (*M* = 402.78, *SD* = 72.33), $r = .714$, $p < .001$, $n = 435$. The winter administration resulted in slightly stronger correlations (*M* = 130.29, *SD* = 40.43) and Alaska SBA (*M* = 402.78, *SD* = 72.33), $r = .718$, $p < .001$, $n = 435$. The strongest correlation was found in the spring administration (*M* = 142.14, *SD* = 42.24) and Alaska SBA (*M* = 402.78, *SD* = 72.33), $r = .719$, $p < .001$, $n = 435$.

Table 6

Correlations Between Fourth Grade Assessment Administrations

3rd Grade Assessment	SBA Reading SS	Fall R-CBM	Winter R-CBM	Spring R-CBM
SBA Reading SS	1	.714**	.718**	.719**

Note. $n = 435$. ** Correlation is significant at the 0.01 level (1-tailed).

Grade 5 results demonstrated that all correlation coefficients were significant at $p < .01$ for each administration of R-CBM (Table 7). There was a statistically significant positive correlation between the fall R-CBM administration (*M* = 128.16, *SD* = 41.97) and Alaska SBA (*M* = 397.24, *SD* = 6.33), $r = .706$, $p < .001$, $n = 517$. The winter administration resulted in slightly stronger correlations (*M* = 146.48, *SD* = 41.832) and Alaska SBA (*M* = 397.24, *SD* = 63.33) , $r = .712$, $p < .001$, $n = 517$. The strongest correlations was found in the spring administration (*M* = 160.96, *SD* = 40.31) and Alaska SBA (*M* = 397.24, *SD* = 63.33) , $r = .717$, $p < .001$, $n = 517$. All correlations were found to be statistically significant at all grades and between each administration of R-CBM. In all cases, the relationship grew progressively stronger between fall, winter, and spring administrations.

Table 7

Correlations Between Fifth Grade Assessment Administrations

5th Grade Assessment	SBA Reading SS	Fall R-CBM	Winter R-CBM	Spring R-CBM
SBA Reading SS	1	.706**	.712**	.717**

Note. $n = 517$. ** Correlation is significant at the 0.01 level (1-tailed).

The second research question of this study addressed to what extent, if at all, could cut scores be derived for each of the three R-CBM testing windows in the fall, winter, and spring to predict success on the Alaska SBA administered in the spring of the same year in the SSD. A binary logistic regression analysis was completed for Grades 3 through 5 with R-CBM score defined as the independent variable and the Alaska SBA score defined as the dependent variable. Outcomes for each grade were reported separately. A binary logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the fall R-CBM.

Logistic Regression Analysis for Grade 3

Grade 3 fall. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the fall R-CBM. Table 8 displays the results of the logistic regression model predicting whether third grade students would pass the Alaska SBA based on their fall R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 472) = 79.527, p < .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 8

Logistic Regression Model for Grade 3 Fall R-CBM as Predictor of Passing Alaska SBA

							95% Confidence Interval for Odd Ratios	
	B	*SE*	Wald	*df*	*p*	Odds Ratio	Lower Limit	Upper Limit
Fall R-CBM	.051	.007	46.285	1	.000	1.052	1.037	1.068
Constant	-.875	.393	4.950	1	.026	.417		

Note. $\chi^2(1, n = 472) = 79.53, p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .339. Predicted success was adequate, with an overall classification of passers and failers, 92.8% classified correctly (Table 9). Specifically, 429 out of 463 students were correctly classified as passing the Alaska SBA (92.66%, true negatives or negative predictive power) and 9 out of 9 students were correctly classified failing the Alaska SBA (100%, true positives or positive predictive power).

Table 9

The Observed and Predicted Frequencies Using Fall R-CBM for Third Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 3rd Grade SBA Proficiency		
Observed 3rd Grade SBA Proficiency	No	Yes	% Correct
No	9	34	20.9
Yes	0	429	100.0
Overall Percentage			92.8

Note. Sensitivity = [9 / (9 + 34)] 100 = 20.93%. Specificity = [429/ (429 + 0)] 100 = 100%. Positive Predictive Power = [9 / (9+0)] 100 = 100%. Negative Predictive Power = [429 / (429 + 34)] 100 = 92.66%.

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 45 at 80% probability. Specifically, a third grade student attaining the cut score of 45 on the R-CBM in the fall is predicted with 80% probability to pass the Alaska SBA. See Figure 8 for a visual representation.

An evaluation of the Grade 3 fall logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. An ROC curve can be constructed by plotting the sensitivity and specificity of each cut off score. Values for sensitivity were plotted on the y axis, and one minus specificity values were plotted on the *x* axis (Figure 9).

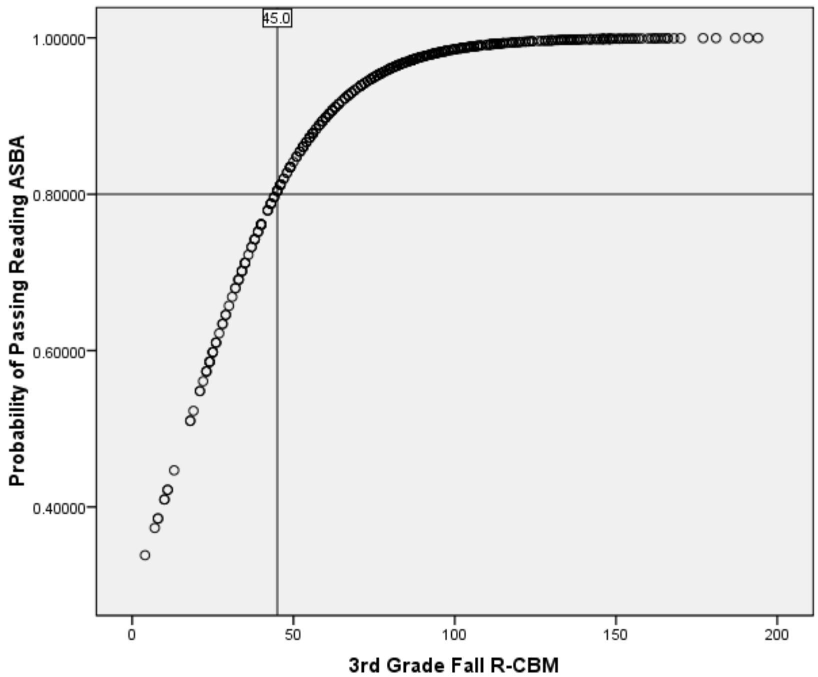


Figure 8. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 3 fall R-CBM to set cut scores.

Swets (1998) suggested that in order to achieve a balance between sensitivity and specificity, the AUC should be greater than .75; however, Minitab (2010) provided a range of acceptability for AUC values. Values closer to 1.0 offered outstanding discrimination; values between .8 and .9 offered excellent discrimination; values between .7 and .8 offered acceptable discrimination, values closer to .6 offered questionable discrimination; but a value of .5 indicated that the predictor variable utility was no better than chance (Minitab, 2010). The resulting area under the curve (AUC) of .856 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 10).

Table 10

Predicted Probability Grade 3 Fall via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.856	.030	.000	.798	.914

Note. Predicted Probability Grade 3 fall has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5

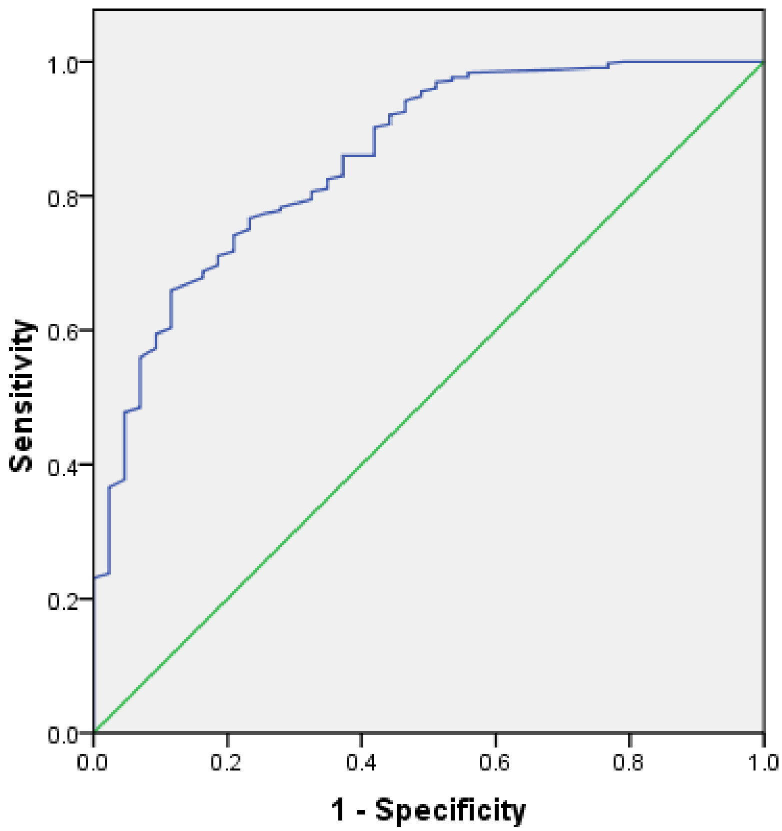


Figure 9. ROC curve for third grade's fall R-CMB.

Grade 3 winter. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the winter R-CBM. Table 11 displays the results of the logistic regression model predicting whether third grade students would pass the Alaska SBA based on their fall R-CBM scores. A test of the full model with R-

CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 472) = 96.226, p < .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 11

Logistic Regression Model for Grade 3 Winter R-CBM as Predictor of Passing Alaska SBA

							95% C.I. for Odds Ratio	
Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	Lower	Upper
R-CBM Grade 3	.051	.007	56.944	1	.000	1.053	1.039	1.067
Constant	-2.213	.524	17.856	1	.000	.109		

$\chi^2(1, n = 472) = 92.226, p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .404. Predicted success was adequate, with an overall classification of passers and failers, 93.6% classified correctly (Table 12). Specifically, 426 out of 453 students were correctly classified as passing the Alaska SBA (94.04%, true negatives or negative predictive power) and 16 out of 19 students were correctly classified failing the Alaska SBA (84.21%, true positives or positive predictive power).

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 70 at 80% probability. Specifically, a third grade student attaining the cut score of 70 on the R-CBM in the winter was predicted with 80% probability to pass the Alaska SBA. See Figure 10 for a visual representation.

Table 12

The Observed and Predicted Frequencies Using Winter R-CBM for Third Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 3rd Grade SBA Proficiency		
Observed 3rd Grade SBA Proficiency	No	Yes	% Correct
No	16	27	37.2
Yes	3	426	99.3
Overall Percentage			93.6

Note. Sensitivity = [16 / (16 + 27)] 100 = 37.21%. Specificity = [426/ (426 + 27)] 100 = 99.3%. Positive Predictive Power = [16 / (16 + 3)] 100 = 84.21%. Negative Predictive Power = [426 / (426 + 27)] 100 = 94.04%.

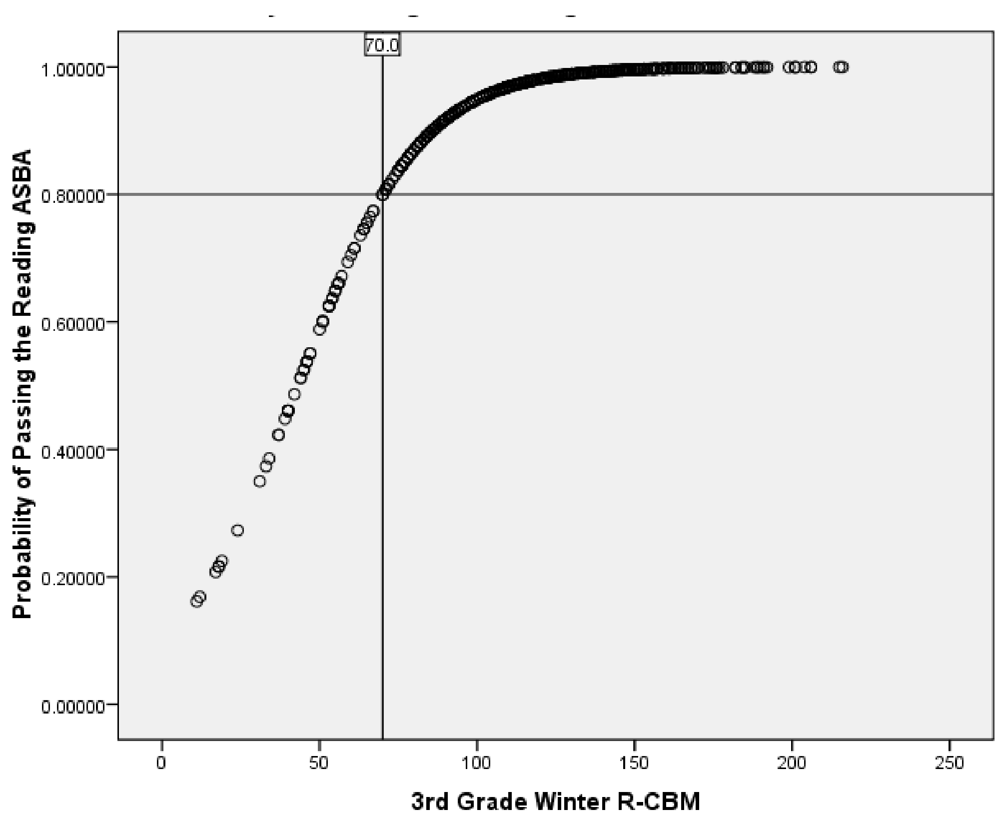


Figure 10. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 3 winter R-CBM to set cut scores.

An evaluation of the Grade 3 winter logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis, and one minus specificity values are plotted on the *x* axis (Figure 11). The resulting AUC of .877 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 13).

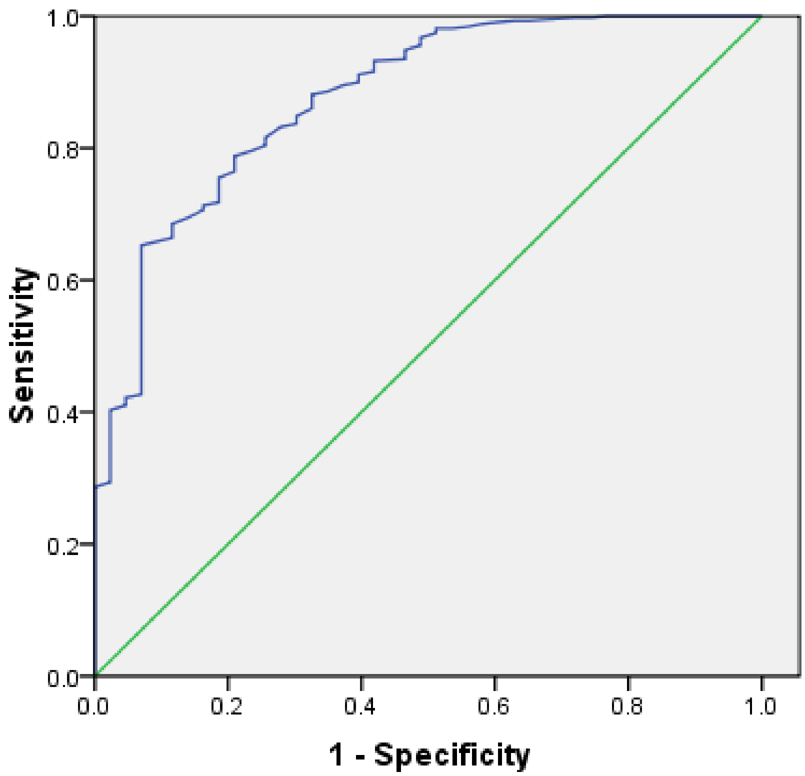


Figure 11. ROC curve for third grade's winter R-CMB.

Table 13

Predicted Probability Grade 3 Winter via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.877	.028	.000	.822	.931

Note. Predicted probability Grade 3 winter has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5

Grade 3 spring. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the spring R-CBM. Table 14 displays the results of the logistic regression model predicting whether third grade students would pass the Alaska SBA based on their spring R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 472) = 113.289$, $p < .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 14

Logistic Regression Model for Grade 3 Spring R-CBM as Predictor of Passing Alaska SBA

Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
R-CBM Grade 3	.056	.007	59.520	1	.000	1.058	1.043	1.073
Constant	-3.326	.644	26.684	1	.000	.036		

Note. *CI* = confidence interval; *B* = intercept; *SE* = standard error; Wald = Wald χ^2 significance; *df* = degree of freedom; UL = upper limit; LL = lower limit. $\chi^2(1, n = 472) = 113.289$, $p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .467. Predicted success was adequate, with an overall classification of passers and failers, 93.9% classified correctly. Specifically, 423 out of 446 students were correctly classified as passing the Alaska SBA (94.84%, true negatives or negative predictive power), and 20 out of 26 students were correctly classified failing the Alaska SBA (76.92%, true positives; Table 15).

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 85 at 80% probability. Specifically, a third grade student attaining the cut score of 85 on the R-CBM in the spring was predicted with 80% probability to pass the Alaska SBA. See Figure 12 for a visual representation.

Table 15

The Observed and Predicted Frequencies Using Spring R-CBM for Third Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 3rd Grade SBA Proficiency		
Observed 3rd Grade SBA Proficiency	No	Yes	% Correct
No	20	23	46.5
Yes	6	423	98.6
Overall Percentage			93.9

Note. Sensitivity = [20 / (20 + 23)] 100 = 46.51%. Specificity = [423/ (423 + 6)] 100 = 98.60%. Positive Predictive Power = [20 / (20 + 6)] 100 = 76.92%. Negative Predictive Power = [423 / (423 + 23)] 100 = 94.84%.

An evaluation of the Grade 3 spring logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 13). The resulting AUC of .900 indicated that the model did an outstanding job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 16).

Table 16

Predicted Probability Grade 3 Spring via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.900	.023	.000	.855	.946

Note. Predicted probability Grade 3 spring has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

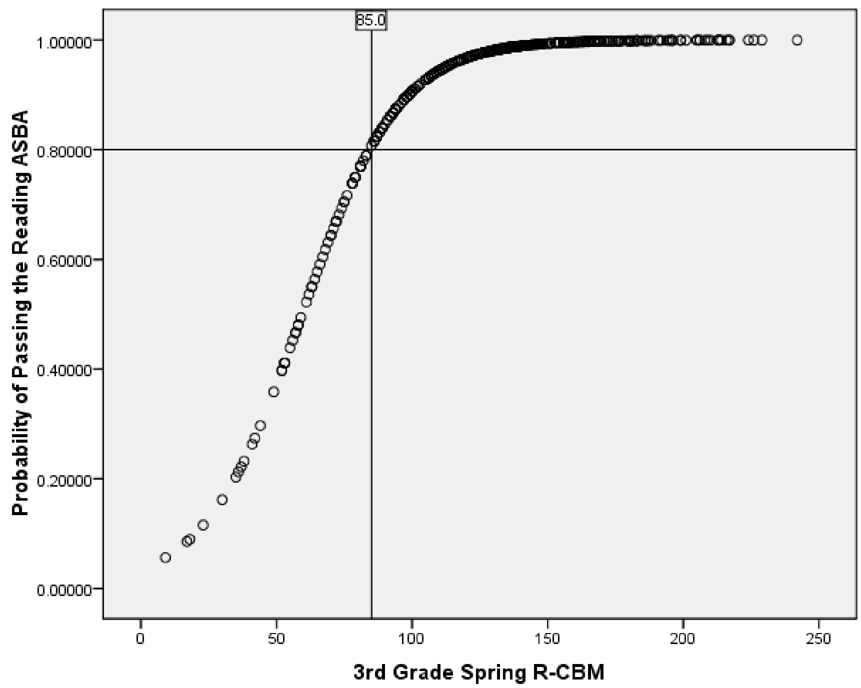


Figure 12. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 3 spring R-CBM to set cut scores.

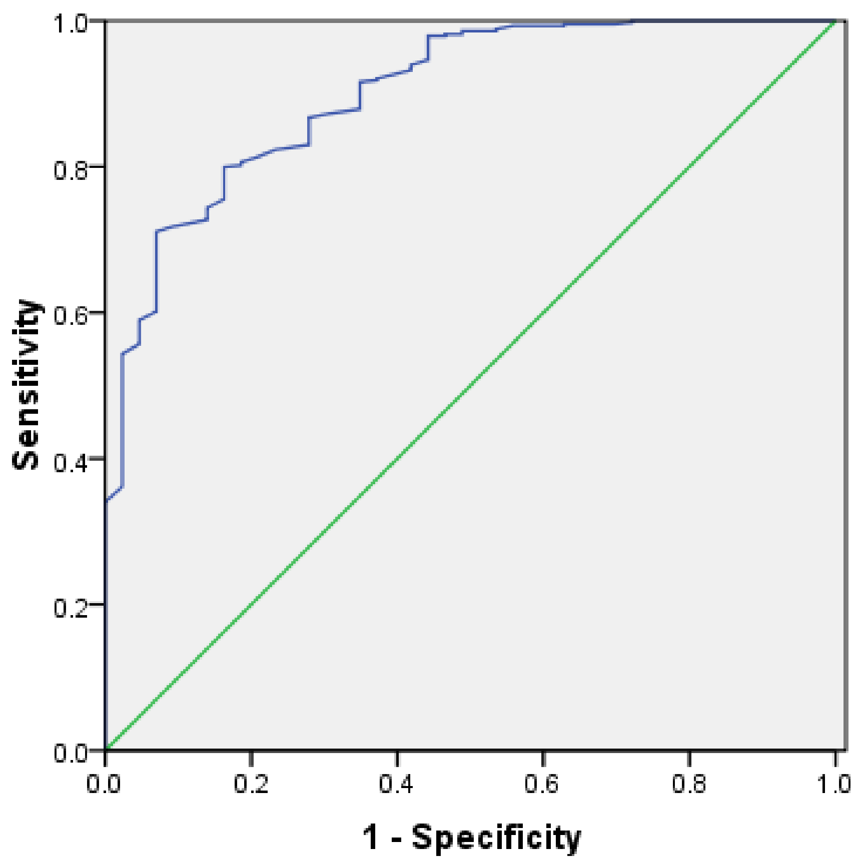


Figure 13. ROC curve for third grade's spring R-CMB.

Logistic Regression Analysis for Grade 4

Grade 4 fall. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the fall R-CBM. Table 17 displays the results of the logistic regression model predicting whether fourth grade students would pass the Alaska SBA based on their fall R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 435) = 131.590$, $p = .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 17

Logistic Regression Model for Grade 4 Fall R-CBM as Predictor of Passing Alaska SBA

Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	95% C.I. for Odds Ratio Lower	Upper
R-CBM Grade 4	.076	.010	56.264	1	.000	1.079	1.058	1.100
Constant	-3.722	.694	28.758	1	.000	.024		

Note. $\chi^2(1, n = 435) = 131.590$, $p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .562. Predicted success was adequate, with an overall classification of passers and failers, 93.8% classified correctly. Specifically, 387 out of 407 students were correctly classified as passing the Alaska SBA (95.09%, true negatives or negative predictive power) and 7 out of 28 students were correctly classified failing the Alaska SBA (75.00%, true positives or positive predictive power; Table 18).

Table 18

The Observed and Predicted Frequencies Using Fall R-CBM for Fourth Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 4th Grade SBA Proficiency		
Observed 4th Grade SBA Proficiency	No	Yes	% Correct
No	21	20	51.2
Yes	7	387	98.2
Overall Percentage			93.8

Note. Sensitivity = [21 / (21 + 20)] 100 = 51.22%. Specificity = [387/ (387 + 7)] 100 = 98.22%. Positive Predictive Power = [21 / (21 + 7)] 100 = 75.00%. Negative Predictive Power = [387 / (387 + 20)] 100 = 95.09%.

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 68 at 80% probability. Specifically, a fourth grade student attaining the cut score of 68 on the R-CBM in the fall is predicted with 80% probability to pass the Alaska SBA. See Figure 14 for a visual representation.

An evaluation of the Grade 4 fall logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. An ROC curve can be constructed by plotting the sensitivity and specificity of each cut off score. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 15). Swets (1998) suggested that in order to achieve a balance between sensitivity and specificity, the AUC should be greater than .75; however, Minitab (2010) provided a range of acceptability for AUC values. Values closer to 1.0 offered outstanding discrimination; values between .8 and .9 offered excellent discrimination; values between .7 and .8 offered acceptable discrimination, values closer to .6 offered questionable discrimination; but a value of .5 indicated that the predictor variable utility was no better than chance (Minitab, 2010). The resulting AUC of .935 indicated that the model did an outstanding job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 19).

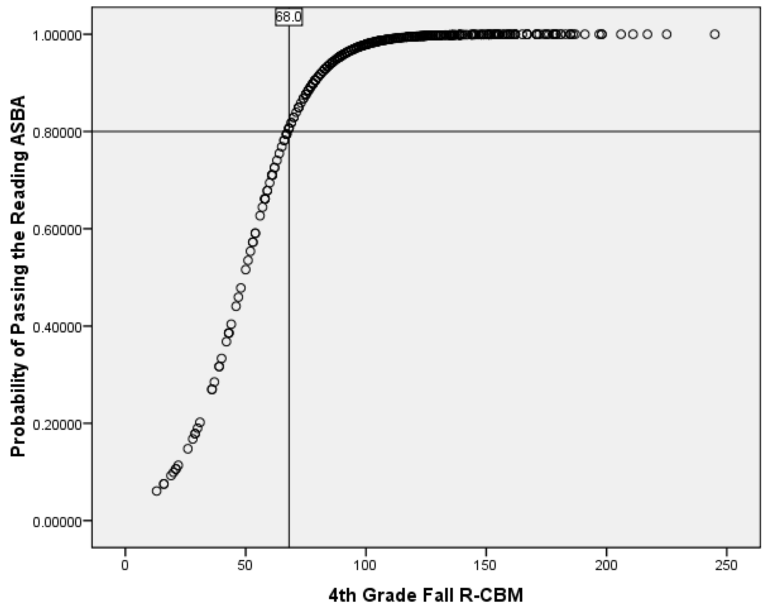


Figure 14. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 4 fall R-CBM to set cut scores.

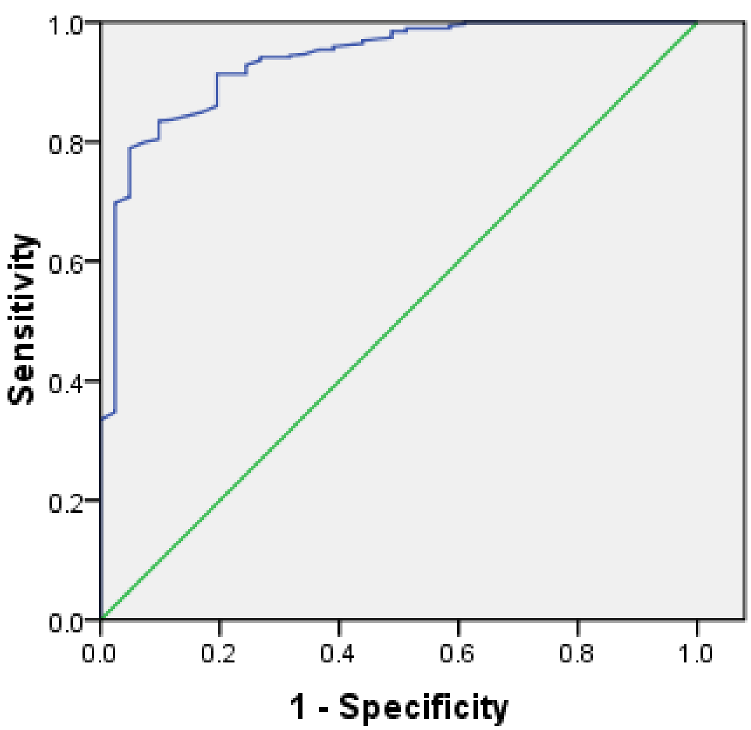


Figure 15. ROC curve for fourth grade spring R-CMB.

Table 19

Predicted Probability Fourth Grade Fall via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.935	.019	.000	.897	.973

Note. Predicted probability Grade 4 fall has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

Grade 4 winter. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the winter R-CBM. Table 20 displays the results of the logistic regression model predicting whether fourth grade students would pass the Alaska SBA based on their winter R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, χ^2 (1, n = 435) = 132.321, p = .001, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 20

Logistic Regression Model for Grade 4 Winter R-CBM as Predictor of Passing Alaska SBA

							95% C.I. for Odds Ratio	
Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	Lower	Upper
R-CBM Grade 4	.071	.009	57.780	1	.000	1.073	1.054	1.093
Constant	-4.867	.843	33.355	1	.000	.008		

Note. χ^2(1, n = 435) = 132.321, p = .001.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .565. Predicted success was adequate, with an overall classification of passers and failers, 93.3% classified correctly. Specifically, 386 out of 407 students were correctly classified as passing the Alaska SBA (94.84%, true negatives or negative predictive power), and 20 out of 28 students were correctly classified failing the Alaska SBA (71.43%, true positives or positive predictive power; Table 21).

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 89 at 80% probability. Specifically, a fourth grade student attaining the cut score of 89 on the R-CBM in the winter is predicted with 80% probability to pass the Alaska SBA. See Figure 16 for a visual representation.

Table 21

The Observed and Predicted Frequencies Using Winter R-CBM for Fourth Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 4th Grade SBA Proficiency		
Observed 4th Grade SBA Proficiency	No	Yes	% Correct
No	20	21	48.8
Yes	8	386	98.0
Overall Percentage			93.3

Note. Sensitivity = [20 / (20 + 21)] 100 = 48.78%. Specificity = [386/ (386 + 8)] 100 = 97.97%. Positive Predictive Power = [20 / (20 + 8)] 100 = 71.43%. Negative Predictive Power = [386 / (386 + 8)] 100 = 94.84%.

An evaluation of the Grade 4 winter logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 17). The resulting AUC of .935 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 22).

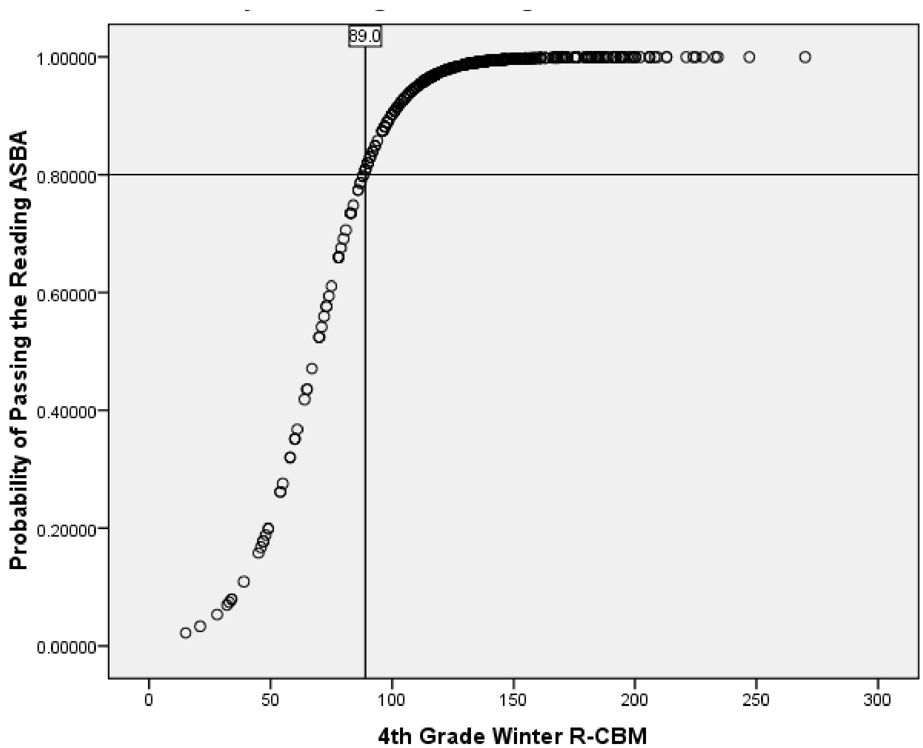


Figure 16. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 4 winter R-CBM to set cut scores.

Table 22

Predicted Probability Fourth Grade Winter via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.935	.019	.000	.897	.973

Note. Predicted probability Grade 4 winter has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

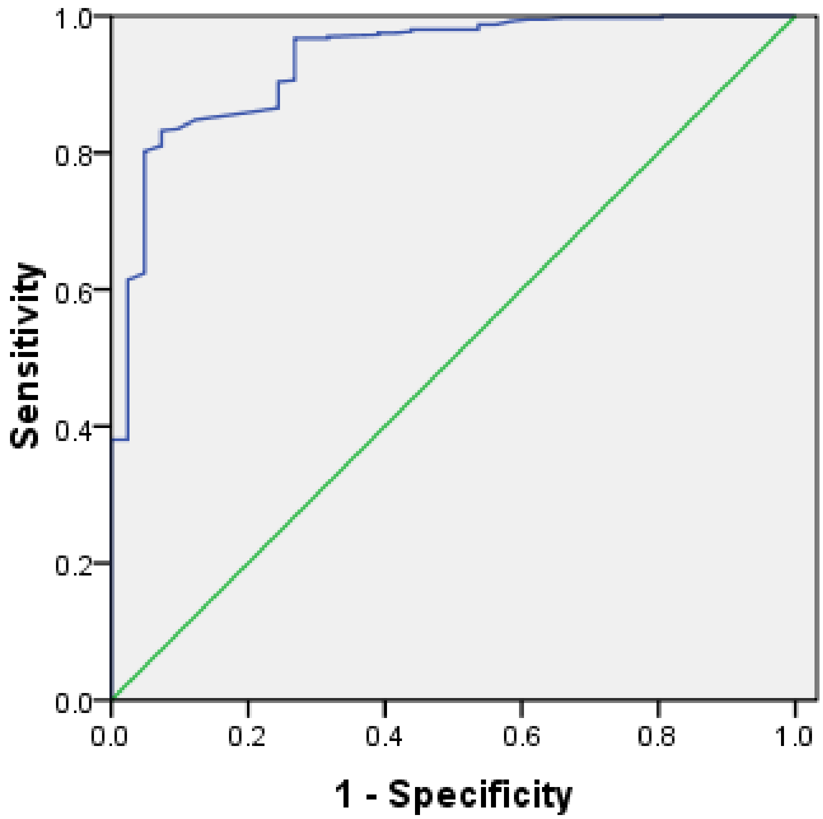


Figure 17. ROC curve for fourth grade winter R-CMB.

Grade 4 spring. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the spring R-CBM. Table 23 displays the results of the logistic regression model predicting whether Grade 4 students would pass the Alaska SBA based on their fall R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 435) = 138.107$, $p = .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 23

Logistic Regression Model for Grade 4 Spring R-CBM as Predictor of Passing Alaska SBA

							95% C.I. for Odds Ratio	
Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	Lower	Upper
R-CBM Grade 4	.075	.011	50.420	1	.000	1.078	1.056	1.101
Constant	-5.916	1.026	33.230	1	.000	.003		

Note. $\chi^2(1, n = 435) = 138.107, p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .586. Predicted success was adequate, with an overall classification of passers and failers, 93.3% classified correctly. Specifically, 387 out of 409 students were correctly classified as passing the Alaska SBA (94.62%, true negatives or negative predictive power) and 19 out of 26 students were correctly classified failing the Alaska SBA (73.08%, true positives or positive predictive power; Table 24).

Table 24

The Observed and Predicted Frequencies Using Spring R-CBM for Fourth Grade Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 4th Grade SBA Proficiency		
Observed 4th Grade SBA Proficiency	No	Yes	% Correct
No	19	22	46.3
Yes	7	387	98.2
Overall Percentage			93.3

Note. Sensitivity = [19 / (19 + 22)] 100 = 46.34%. Specificity = [387/ (387 + 7)] 100 = 98.22%. Positive Predictive Power = [19 / (19 + 7)] 100 = 73.08%. Negative Predictive Power = [387 / (387 + 7)] 100 = 94.62%.

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 98 at 80% probability. Specifically, a Grade 4 student attaining the cut score of 98 on the R-CBM in the spring is predicted with 80% probability to pass the Alaska SBA. See Figure 18 for a visual representation.

An evaluation of the Grade 4 spring logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 19). The resulting AUC of .945 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 25).

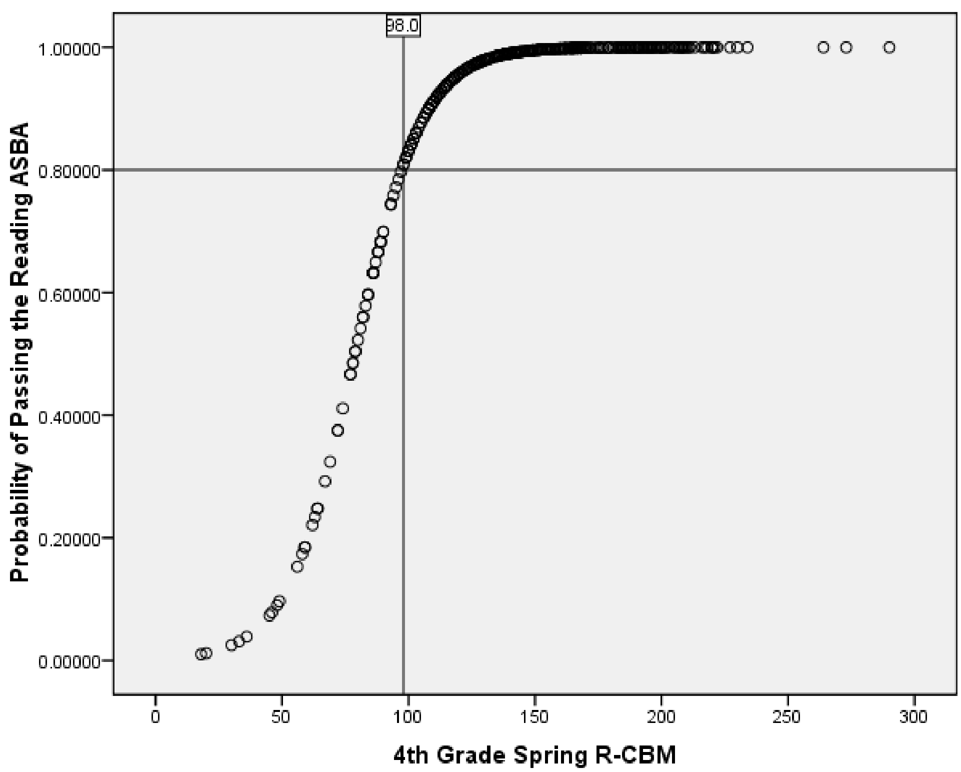


Figure 18. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 4 spring R-CBM to set cut scores.

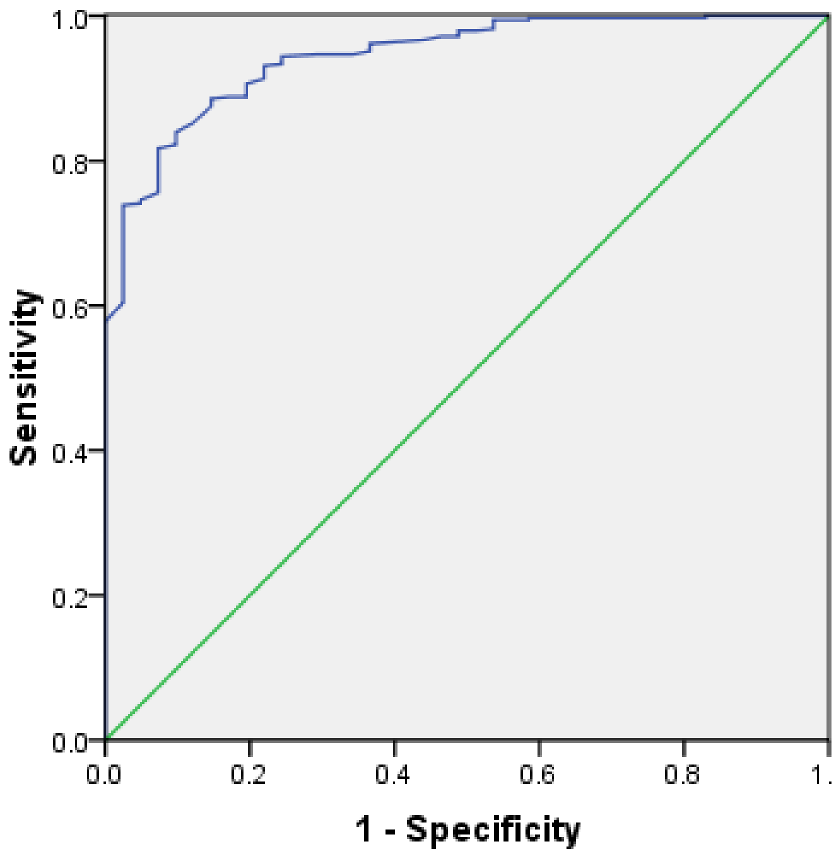


Figure 19. ROC curve for fourth grade spring R-CMB.

Table 25

Predicted Probability Fourth Grade Spring via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.945	.015	.000	.916	.973

Note. Predicted probability Grade 4 spring has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

Logistic Regression Analysis for Grade 5

Grade 5 fall. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the fall R-CBM. Table 26 displays the results of the logistic regression model predicting whether Grade 5 students would pass the Alaska SBA based on their fall R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 517) = 81.407$, $p = .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .407 predicted success was adequate, with an overall classification of passers and failers, 95.9% classified correctly. Specifically, 486 of 506 students were correctly classified as passing the Alaska SBA (96.05%, true negatives or negative predictive power) and 10 of 11 students were correctly classified failing the Alaska SBA (90.91%, true positives or positive predictive power; Table 27).

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 73 at 80% probability. Specifically, a Grade 5 student attaining the cut score of 73 on the R-CBM in the fall is predicted with 80% probability to pass the Alaska SBA. See Figure 20 for a visual representation.

Table 26

Logistic Regression Model for Grade 5 Fall R-CBM as Predictor of Passing Alaska SBA

							95% C.I. for Odds Ratio	
Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	Lower	Upper
R-CBM Grade 4	.058	.009	44.427	1	.000	1.060	1.042	1.078
Constant	-2.831	.729	15.092	1	.000	.059		

Note. $\chi^2(1, n = 517) = 81.407$, $p = .001$.

Table 27

The Observed and Predicted Frequencies Using Fall R-CBM for Grade 5 Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 5th Grade SBA Proficiency		
Observed 5th Grade SBA Proficiency	No	Yes	% Correct
No	10	20	33.3
Yes	1	486	99.8
Overall Percentage			95.9

Note. Sensitivity = [10 / (10 + 20)] 100 = 33.33%. Specificity = [486/ (486 + 1)] 100 = 99.79%. Positive Predictive Power = [10 / (10 + 1)] 100 = 90.91%. Negative Predictive Power = [486 / (486 + 20)] 100 = 96.05%.

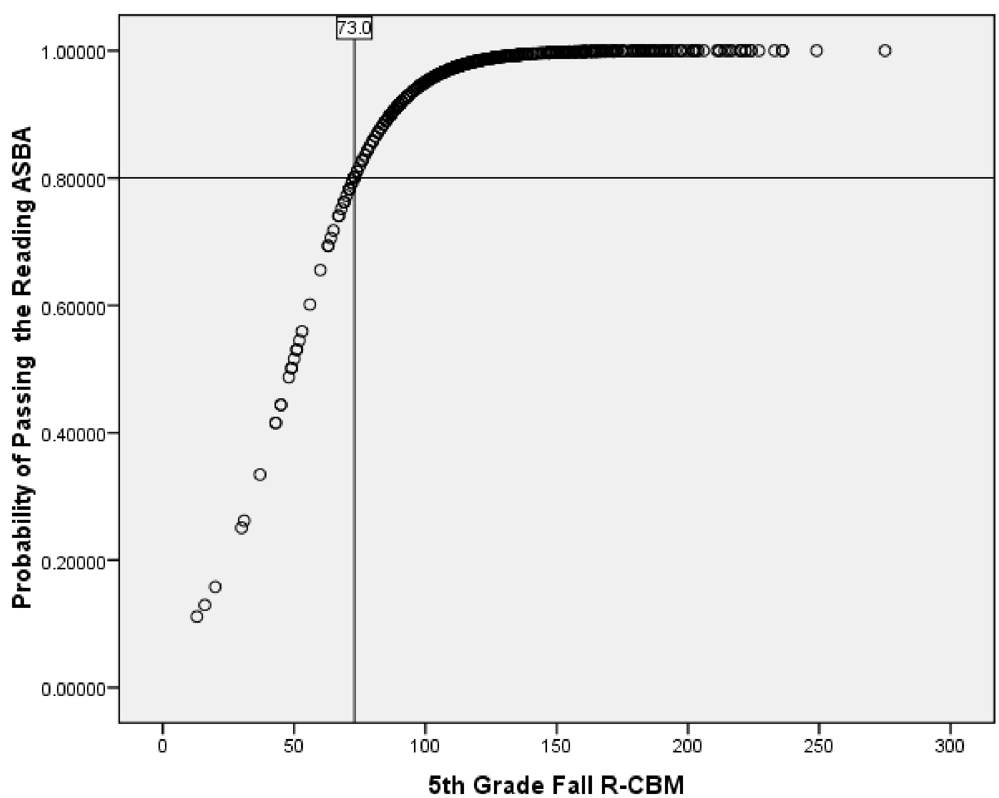


Figure 20. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 5 fall R-CBM to set cut scores.

An evaluation of the Grade 5 fall logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. A ROC curve can be constructed by plotting the sensitivity and specificity of each cut off score. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 21). Swets (1998) suggested that in order to achieve a balance between sensitivity and specificity, the AUC should be greater than .75; however, Minitab (2010) provided a range of acceptability for AUC values. Values closer to 1.0 offered outstanding discrimination; values between .8 and .9 offered excellent discrimination; values between .7 and .8 offered acceptable discrimination, values closer to .6 offered questionable discrimination; but a value of .5 indicated that the predictor variable utility was no better than chance (Minitab, 2010). The resulting AUC of .893 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 28).

Table 28

Predicted Probability Grade 5 Fall via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.893	.030	.000	.834	.951

Note. Predicted Probability Grade 5 fall has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

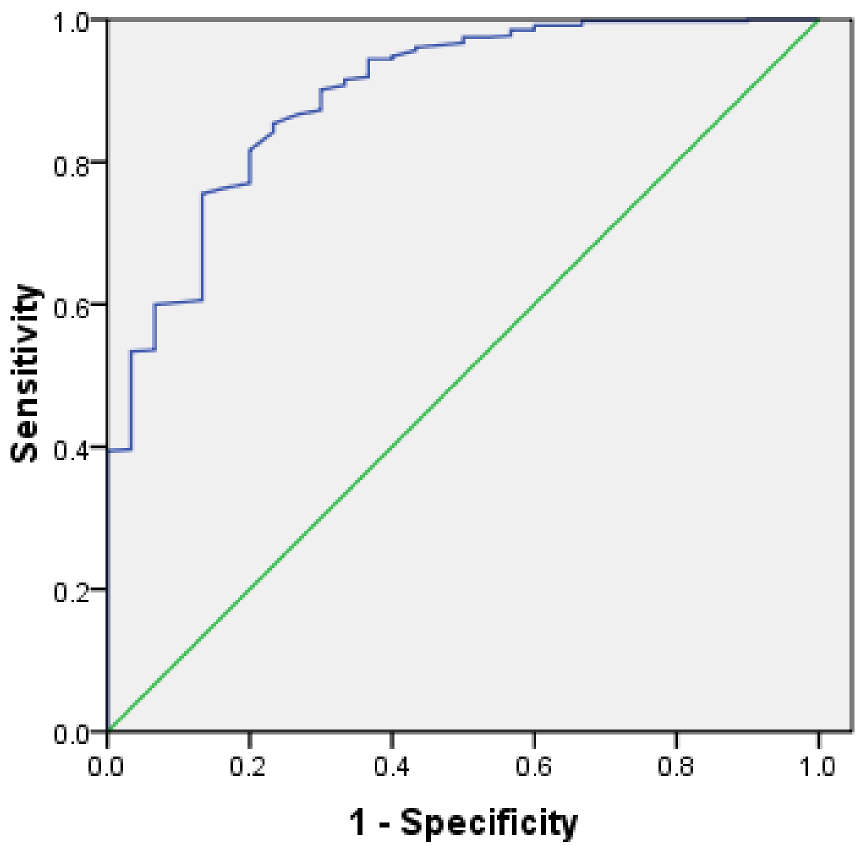


Figure 21. ROC curve for fifth grade fall R-CMB.

Grade 5 winter. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the winter R-CBM. Table 29 displays the results of the logistic regression model predicting whether fifth grade students would pass the Alaska SBA based on their Winter R-CBM scores. A test of the full model with R-CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 517) = 81.406$, $p = .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 29

Logistic Regression Model for Grade 5 Winter R-CBM as Predictor of Passing Alaska SBA

Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	95% C.I. for Odds Ratio Lower	Upper
R-CBM Grade 4	.053	.008	45.552	1	.000	1.054	1.038	1.070
Constant	-3.283	.787	17.403	1	.000	.038		

Note. $\chi^2(1, n = 517) = 81.406, p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .407 predicted success was adequate, with an overall classification of passers and failers, 95.0 classified correctly. Specifically, 484 of 507 students were correctly classified as passing the Alaska SBA (95.46%, true negatives or negative predictive power), and 7 of 10 students were correctly classified failing the Alaska SBA (70.00%, true positives or positive predictive power, Table 30).

Table 30

The Observed and Predicted Frequencies Using Winter R-CBM for Grade 5 Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 5th Grade SBA Proficiency		
Observed 5th Grade SBA Proficiency	No	Yes	% Correct
No	7	23	23.3
Yes	3	484	99.4
Overall Percentage			95.0

Note. Sensitivity = [7 / (7 + 23)] 100 = 23.33%. Specificity = [484/ (484 + 3)] 100 = 99.38%. Positive Predictive Power = [7 / (7 + 3)] 100 = 70.00%. Negative Predictive Power = [484 / (484 + 3)] 100 = 95.46%.

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 89 at 80% probability. Specifically, a Grade 5 student attaining the cut score of 89 on the R-CBM in the winter is predicted with 80% probability to pass the Alaska SBA. See Figure 22 for a visual representation.

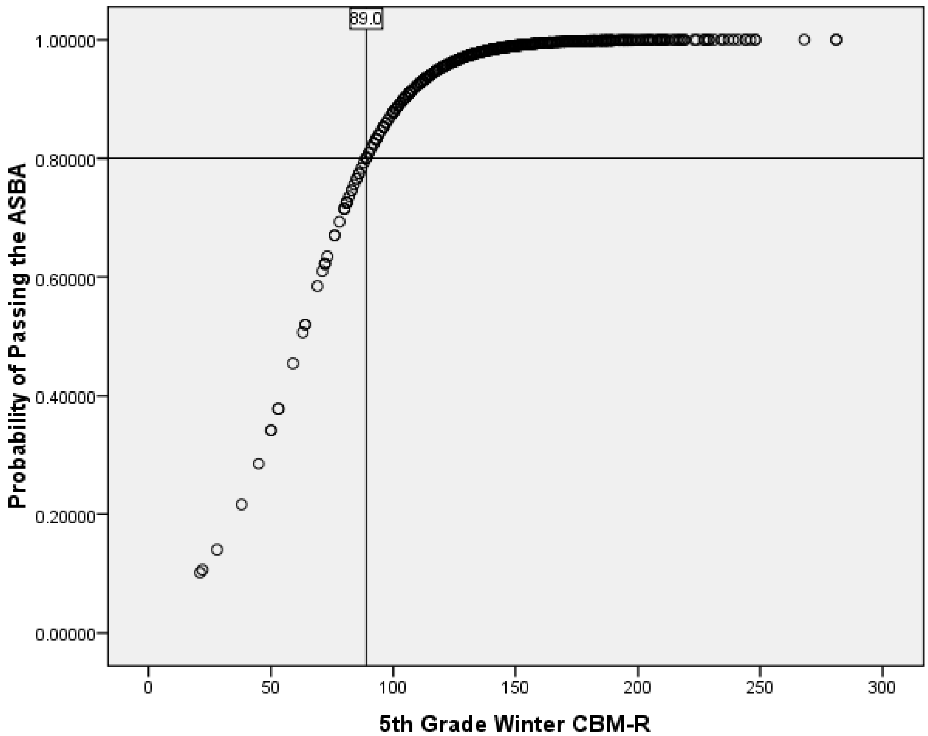


Figure 22. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 5 winter R-CBM to set cut scores.

An evaluation of Grade 5 winter logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 23). The resulting AUC of .903 indicates that the model does an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 31).

Table 31

Predicted Probability Grade 5 Winter via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.903	.025	.000	.853	.952

Note. Predicted probability Grade 5 winter has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

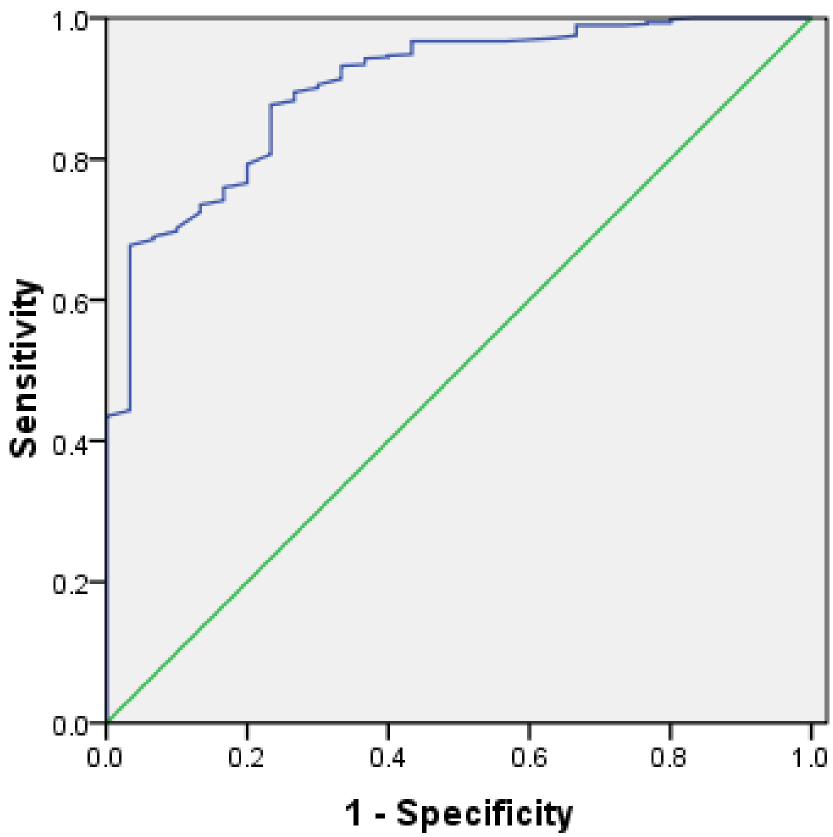


Figure 23. ROC curve for fifth grade winter R-CMB.

Grade 5 spring. A direct logistic regression analysis was performed on failing/passing the Alaska SBA the outcome (coded 0 = failing the Alaska SBA and 1 = passing the Alaska SBA) and one predictor, the spring R-CBM. Table 32 displays the results of the logistic regression model predicting whether Grade 5 students would pass the Alaska SBA based on their spring R-CBM scores. A test of the full model with R-

CBM as predictor against a constant-only model was statistically significant, $\chi^2(1, n = 517) = 93.188$, $p = .001$, indicating that the R-CBM distinguished between passing and failing the Alaska SBA.

Table 32

Logistic Regression Model for Grade 5 Spring R-CBM as Predictor of Passing Alaska SBA

							95% C.I. for Odds Ratio	
Variable	*B*	*SE*	Wald	*df*	*p*	Odds Ratio	Lower	Upper
R-CBM Grade 4	.056	.008	47.191	1	.000	1.057	1.041	1.074
Constant	-4.275	.896	22.744	1	.000	.014		

Note. $\chi^2(1, n = 517) = 93.188$, $p = .001$.

The variance in pass rate accounted for is significant, with Nagelkerke R^2 equal to .461. Predicted success was adequate, with an overall classification of passers and failers, 95.0% classified correctly. Specifically, 482 of 503 students were correctly classified as passing the Alaska SBA (95.83%, true negatives or negative predictive power) and 9 of 14 students were correctly classified failing the Alaska SBA (64.29%, true positives or positive predictive power; Table 33).

A comparison of the means of the predicted probability for each cut score was conducted using the comparison of means function in SPSS. This analysis yielded a cut score of 102 at 80% probability. Specifically, a Grade 5 student attaining the cut score of 102 on the R-CBM in the spring is predicted with 80% probability to pass the Alaska SBA. See Figure 24 for a visual representation.

An evaluation of the Grade 5 spring logistic regression model's ability to classify observations correctly was completed with a receiver operating characteristic (ROC) curve. Values for sensitivity are plotted on the y axis and 1- specificity values are plotted on the *x* axis (Figure 25). The resulting AUC of .921 indicated that the model did an excellent job of predicting an observations response (Hosmer & Lemeshow, 2000; Table 34).

Table 33

The Observed and Predicted Frequencies Using Spring R-CBM for Grade 5 Pass Rates on the Alaska SBA by Logistic Regression with the Predicted Probabilities Cutoff of .5

	Predicted 5th Grade SBA Proficiency		
Observed 5th Grade SBA Proficiency	No	Yes	% Correct
No	9	21	30.0
Yes	5	482	99.0
Overall Percentage			95.0

Note. Sensitivity = [9 / (9 + 21)] 100 = 30.00%. Specificity = [482/ (482 + 5)] 100 = 98.97%. Positive Predictive Power = [9 / (9 + 5)] 100 = 64.29%. Negative Predictive Power = [482 / (482 + 21)] 100 = 95.83%.

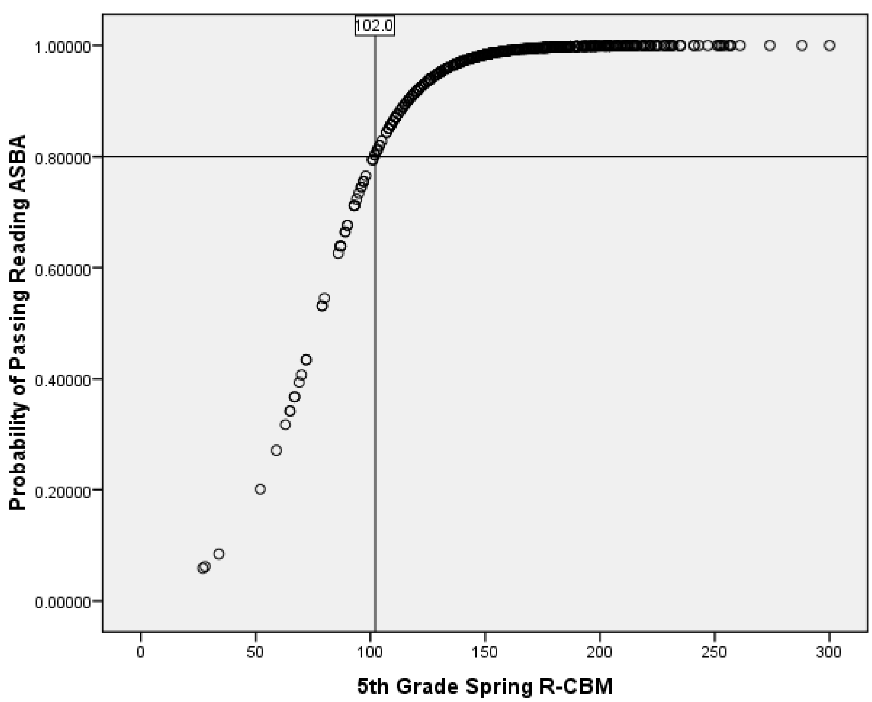


Figure 24. Scatter plot demonstrating predicted probabilities of passing the reading ASBA as a function of the Grade 5 spring R-CBM to set cut scores.

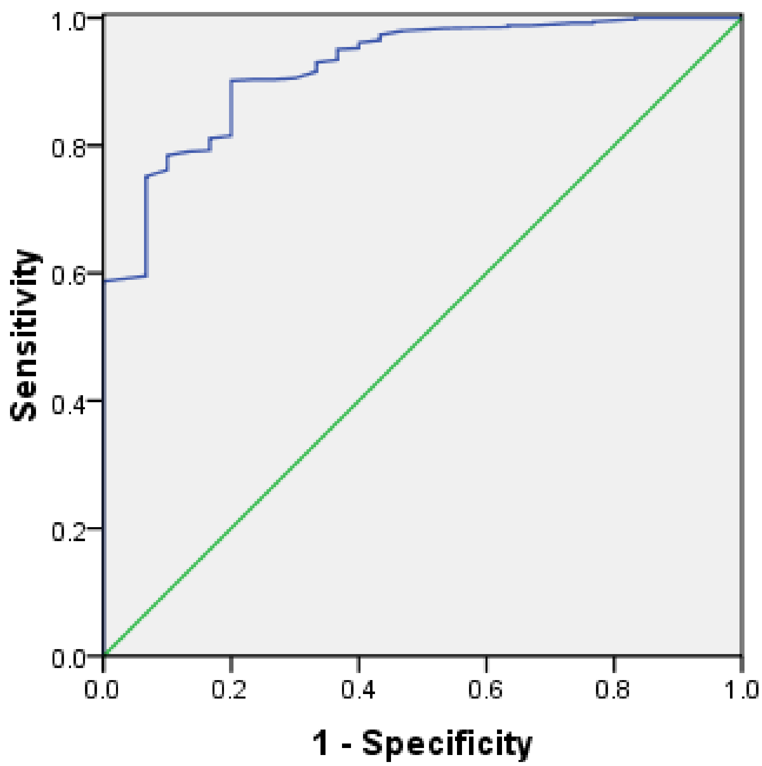


Figure 25. ROC curve for fifth grade spring R-CMB.

Table 34

Predicted Probability Grade 5 Spring via AUC

			Asymptotic 95% Confidence Interval	
AUC	*SE*	*p**	Lower Bound	Upper Bound
.921	.022	.000	.879	.964

Note. Predicted Probability Grade 5 spring has at least one tie between the positive actual state group and the negative actual state group. Statistics might be biased. * Null hypothesis: true area = 0.5.

Cross Validation

A logistic regression analysis revealed that statistically derived cut scores could be established and used to identify students that at risk for failing the Alaska SBA. Cross validation of cut scores established in this study was completed using a cross tab analysis. The scores established using the student performance data from fiscal year 2010 (FY10)

Alaska SBA were applied to student outcomes on the fiscal year 2011 (FY11) Alaska SBA. Student scores for the FY11 R-CBM for fall, winter, and spring were dichotomized with scores below the established cut scores were coded 0, scores at or above the established cut score were coded 1. Similarly, student scores on the Alaska SBA were coded 0 for scores that did not meet proficiency and 1 for scores that were equal to or greater than the proficiency level. The cross tab analysis revealed that the statistically derived scores established for FY10 data classified student performance for the FY11 school year with levels adequate for use within an RTI framework. Overall correct classification was approximately 3% lower than the previous year.

Specificity or NPP ranged from approximately 88% to 91% for the cross validation compared to approximately 93% to 95% in the logistic model developed with the FY10 data. There were however more drastic differences with the cross validation with levels of sensitivity. Levels of sensitivity in the model ranged from 64% to 100% compared to the cross validation with levels ranging from 46% to 71%. Chapter 2 discussed common statistics used as a measure of diagnostic accuracy. Scatter plots of student scores for Alaska SBA and R-CBM were used to illustrate each of these diagnostic efficacy statistics for each of the screenings Grades 3 through 5. The scatter plots illustrate sensitivity, specificity, PPP, and NPP for each R-CBM administration relative to same grade Alaska SBA outcomes.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 26 illustrates that a fall CBM-R cut score of 45 established via logistic regression appears to be a valid indicator of student performance on the third grade Alaska SBA administered in the spring of the same school year. A total of 18.76% of the students (94 out of 501) were predicted to fail the Alaska SBA based on the fall R-CBM cut score, yet only 44 out of 94 students, or 46.81%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 81.24% of the students (407 out of 501) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 401 of the 407 students, or 98.53%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 50 students scored less than proficient on the Alaska SBA. Of those, 44 or 88% were accurately predicted by the fall R-CBM cut score. Conversely, 401 out of the 451 students, or 88.91%, passing the Alaska SBA were accurately predicted by the fall R-CBM.

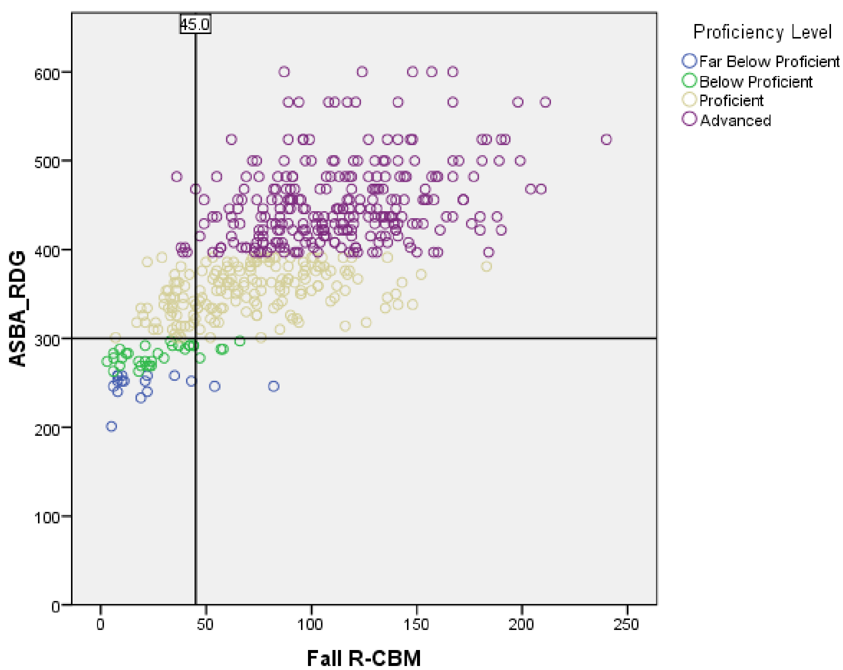


Figure 26. Diagnostic accuracy for Alaska SBA by Grade 3 fall R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 27 illustrates that a winter CBM-R cut score of 70 established via logistic regression appears to be a valid indicator of student performance on the third grade Alaska SBA administered in the spring of the same school year. A total of 16.94% of the students (93 out of 549) were predicted to fail the Alaska SBA based on the winter R-CBM cut score, yet only 48 out of 93 students, or 51.61%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 81.24% of the students (456 out of 549) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 448 of the 456 students, or 98.25%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 56 students scored less than proficient on the Alaska SBA. Of those, 48 or 85.71% were accurately predicted by the winter R-CBM cut score. Conversely, 401 out of 451 students, or 88.91%, passing the Alaska SBA were accurately predicted by the winter R-CBM.

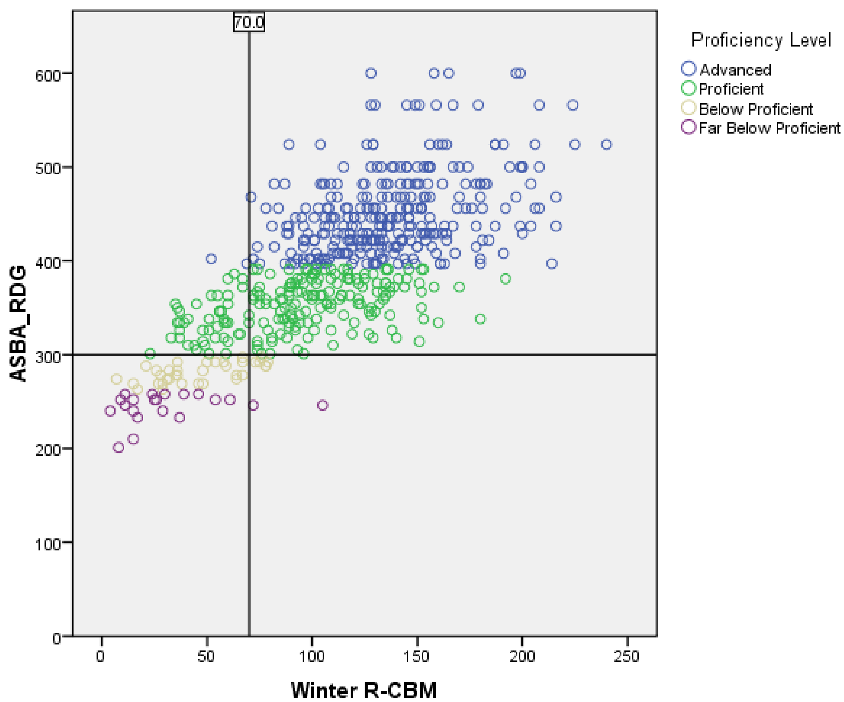


Figure 27. Diagnostic accuracy for Alaska SBA by Grade 3 winter R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 28 illustrates that a spring CBM-R cut score of 85 established via logistic regression appears to be a valid indicator of student performance on the third grade Alaska SBA administered in the spring of the same school year. A total of 16.23% of the students (92 out of 567) were predicted to fail the Alaska SBA based on the spring R-CBM cut score, yet only 51 out of 92 students, or 55.43%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 83.77% of the students (475 out of 567) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 467 out of 475 students, or 98.32%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 59 students scored less than proficient on the Alaska SBA. Of those, 51 or 86.44% were accurately predicted by the spring R-CBM cut score. Conversely, 448 out of 493 students, or 90.87%, passing the Alaska SBA were accurately predicted by the spring R-CBM.

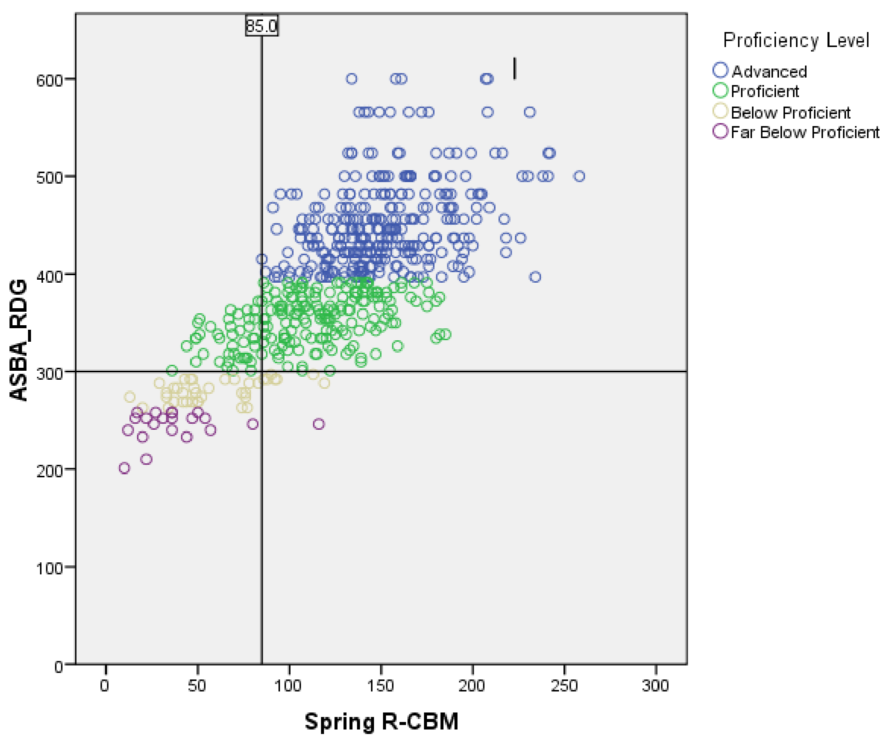


Figure 28. Diagnostic accuracy for Alaska SBA by Grade 3 spring R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 29 illustrates that a fall CBM-R cut score of 68 established via logistic regression appears to be a valid indicator of student performance on the fourth grade Alaska SBA administered in the spring of the same school year. A total of 11.68% of the students (64 out of 548) were predicted to fail the Alaska SBA based on the fall R-CBM cut score, yet only 35 out of 64 students, or 54.69%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 88.32% of the students (484 out of 548) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 460 out of 484 students, or 95.04%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 59 students scored less than proficient on the Alaska SBA. Of those, 35 or 59.32% were accurately predicted by the fall R-CBM cut score. Conversely, 460 out of 489 students, or 94.07%, passing the Alaska SBA were accurately predicted by the fall R-CBM.

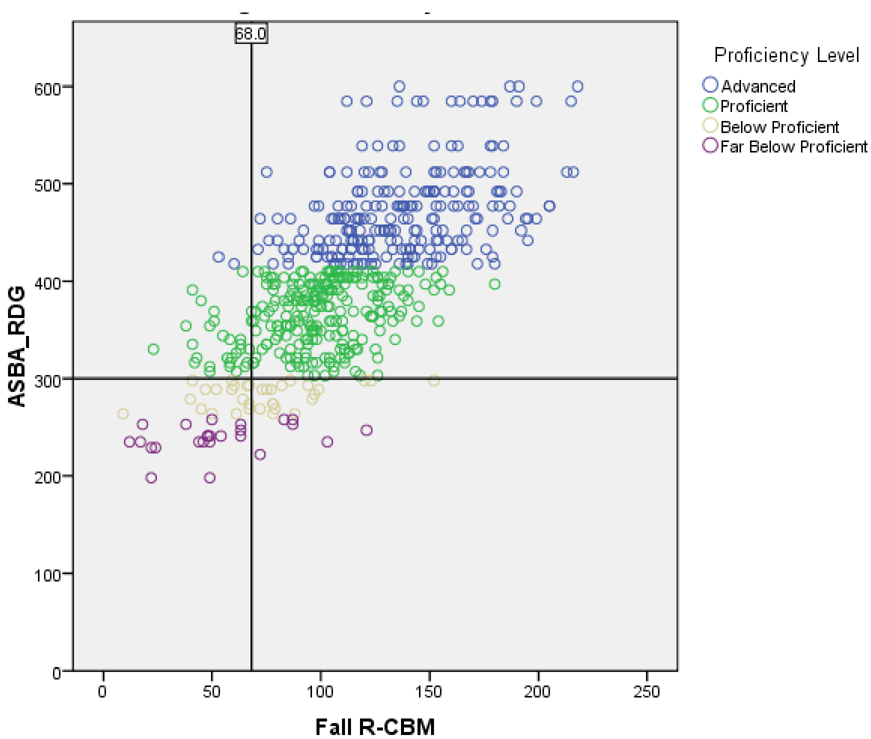


Figure 29. Diagnostic accuracy for Alaska SBA by Grade 4 fall R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 30 illustrates that a winter CBM-R cut score of 89 established via logistic regression appears to be a valid indicator of student performance on the fourth grade Alaska SBA administered in the spring of the same school year. A total of 12.87% of the students (70 out of 544) predicted to fail the Alaska SBA based on the winter R-CBM cut score, yet only 36 out of 70 students, or 51.43%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 81.24% of the students (474 out of 544) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 446 out of 474 students, or 94.09%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 64 students scored less than proficient on the Alaska SBA. Of those, 36 or 56.25% were accurately predicted by the winter R-CBM cut score. Conversely, 446 out of 480 students, or 92.92%, passing the Alaska SBA were accurately predicted by the winter R-CBM.

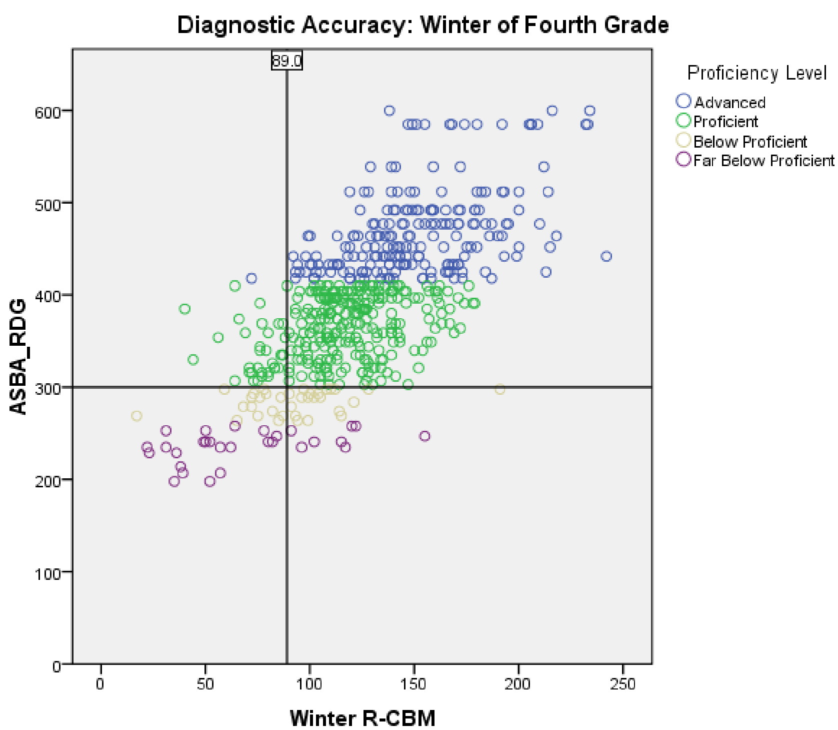


Figure 30. Diagnostic accuracy for Alaska SBA by Grade 4 winter R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 31 illustrates that a spring CBM-R cut score of 98 established via logistic regression appears to be a valid indicator of student performance on the fourth grade Alaska SBA administered in the spring of the same school year. A total of 12.34% of the students (68 out of 551) predicted to fail the Alaska SBA based on the spring R-CBM cut score, yet only 34 out of 68 students, or 50%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 87.66% of the students (483 out of 551) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 452 out of 483 students, or 93.58%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 65 students scored less than proficient on the Alaska SBA. Of those, 34 or 52.31% were accurately predicted by the spring R-CBM cut score. Conversely, 452 out of 486 students, or 93%, passing the Alaska SBA were accurately predicted by the spring R-CBM.

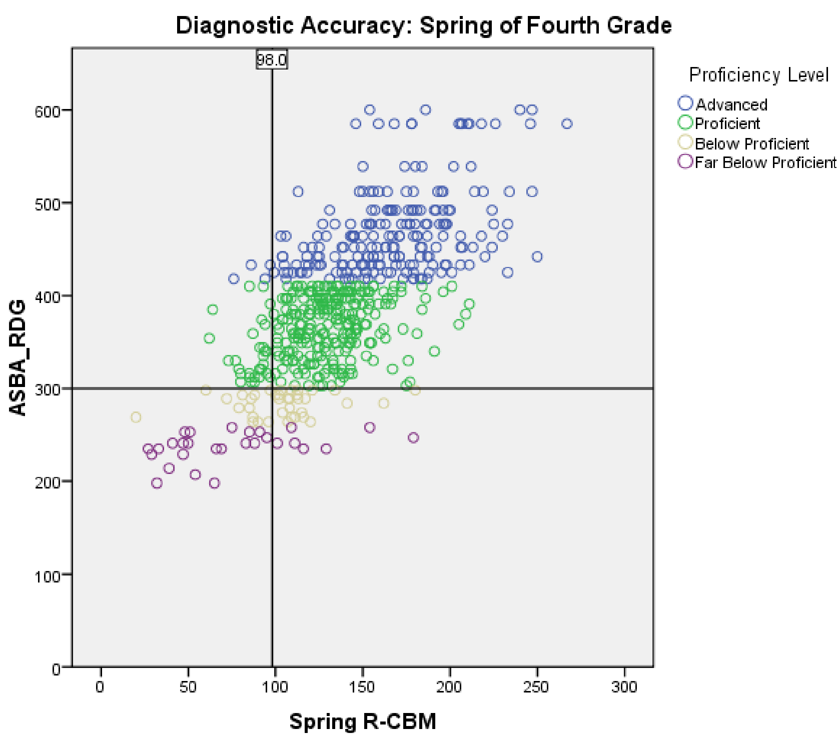


Figure 31. Diagnostic accuracy for Alaska SBA by Grade 4 spring R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 32 illustrates that a fall CBM-R cut score of 73 established via logistic regression appears to be a valid indicator of student performance on the fifth grade Alaska SBA administered in the spring of the same school year. A total of 9.61% of the students (49 out of 510) predicted to fail the Alaska SBA based on the fall R-CBM cut score, yet only 34 out of 49 students, or 69.39%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 90.31% of the students (461 out of 510) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 430 out of 461students, or 93.28%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 65 students scored less than proficient on the Alaska SBA. Of those, 34 or 52.31% were accurately predicted by the fall R-CBM cut score. Conversely, 430 out of 445 students, or 96.63%, passing the Alaska SBA were accurately predicted by the fall R-CBM.

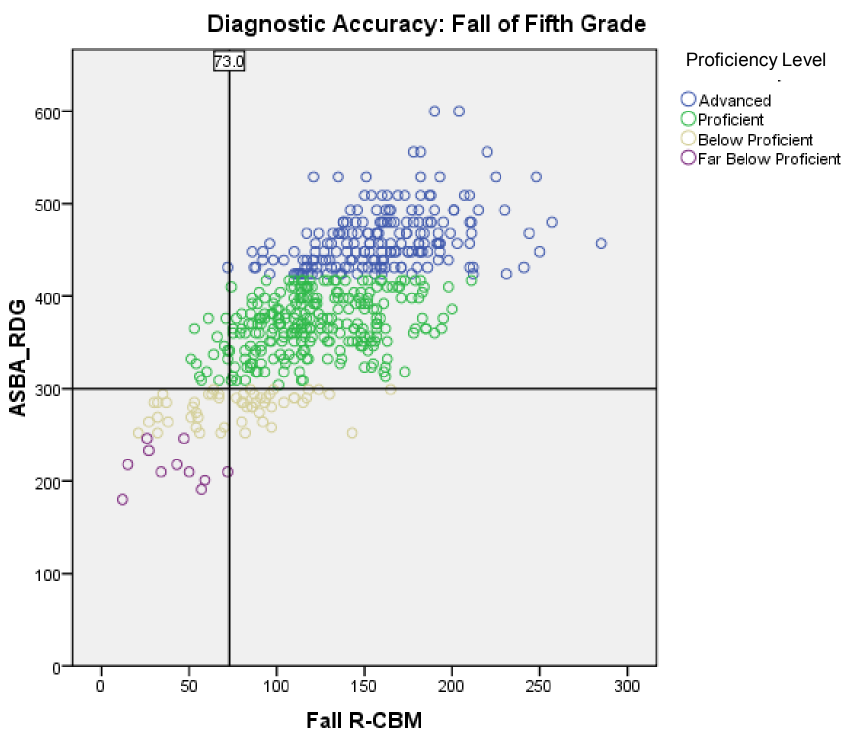


Figure 32. Diagnostic accuracy for Alaska SBA by Grade 5 fall R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 33 illustrates that a winter CBM-R cut score of 89 established via logistic regression appears to be a valid indicator of student performance on the fifth grade Alaska SBA administered in the spring of the same school year. A total of 10.39% of the students (51 out of 491) were predicted to fail the Alaska SBA based on the winter R-CBM cut score, yet only 36 out of 51 students, or 70.59%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 82.89% of the students (440 out of 491) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 403 out of 440 students, or 91.59%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 73 students scored less than proficient on the Alaska SBA. Of those, 36 or 49.32% were accurately predicted by the winter R-CBM cut score. Conversely, 403 out of 418 students, or 96.41%, passing the Alaska SBA were accurately predicted by the winter R-CBM.

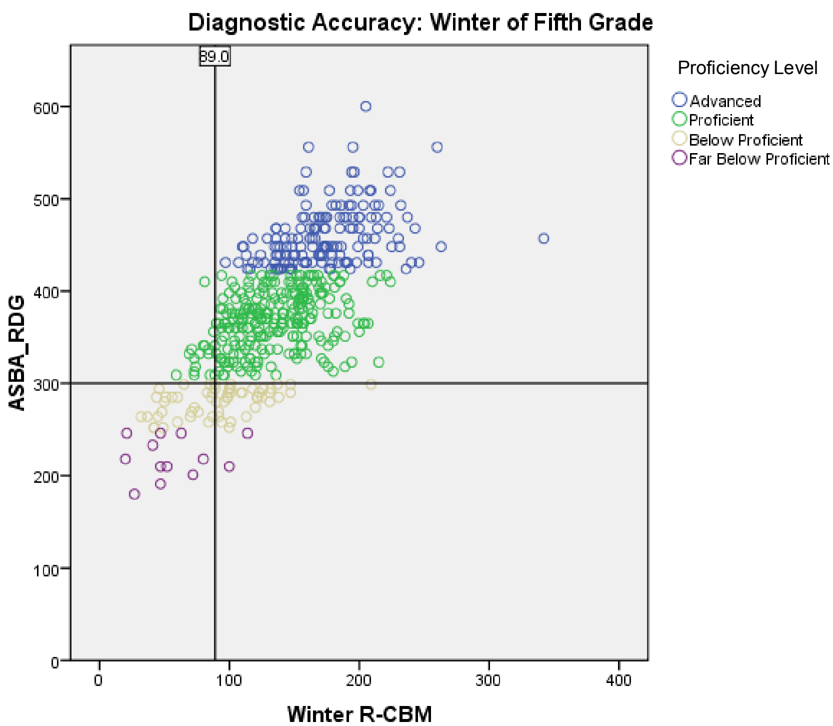


Figure 33. Diagnostic accuracy for Alaska SBA by Grade 5 winter R-CBM.

Scores established in the study using FY10 data were applied to the FY11 data. Figure 34 illustrates that a spring CBM-R cut score of 102 established via logistic regression appears to be a valid indicator of student performance on the fifth grade Alaska SBA administered in the spring of the same school year. A total of 13.24% of the students (65 out of 491) were predicted to fail the Alaska SBA based on the spring R-CBM cut score, yet only 41 out of 65 students, or 63.08%, predicted to fail scored below proficient on the Alaska SBA (i.e., positive predictive power). In comparison, 86.76% of the students (426 out of 491) were predicted to pass the Alaska SBA based on the fall R-CBM cut score. Of these, 397 out of 426 students, or 93.19%, scored proficient or advanced on the Alaska SBA (i.e., negative predictive power). Overall, 70 students scored less than proficient on the Alaska SBA. Of those, 41 or 58.57% were accurately predicted by the fall R-CBM cut score. Conversely, 397 out of 421 students, or 94.3%, passing the Alaska SBA were accurately predicted by the fall R-CBM.

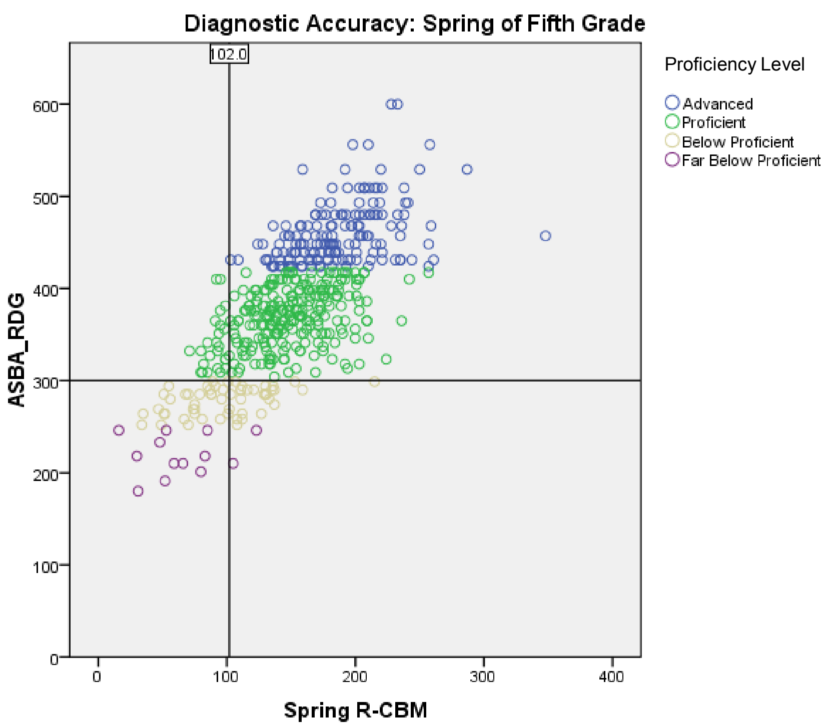


Figure 34. Diagnostic accuracy for Alaska SBA by Grade 5 spring R-CBM.

The overall model cross validation summary for FY10 is shown in Table 35. The overall model cross validation summary for FY11 can be reviewed in Table 36.

Table 35

FY10 Logistic Regression Model Summary

R-CBM	*n*	True Positive	False Positive	True Negative	False Negative	% Positive Predictive Power	% Negative Predictive Power	% Overall Correct Classification	% Sensitivity	% Specificity
Grade 3 Fall	472	9	0	429	34	100.00	92.66	92.80	20.93	100.00
Grade 3 Winter	472	16	3	426	27	84.21	94.04	93.64	37.21	99.30
Grade 3 Spring	472	20	6	423	26	76.92	94.21	93.26	46.51	94.84
Grade 4 Fall	435	21	7	387	20	75.00	95.09	93.79	51.22	98.22
Grade 4 Winter	435	20	8	386	21	71.43	94.84	93.33	48.78	97.97
Grade 4 Spring	435	19	7	387	22	73.08	94.62	93.33	46.34	98.22
Grade 5 Fall	517	10	1	486	20	90.91	96.05	95.94	33.33	99.79
Grade 5 Winter	517	7	3	484	23	70.00	95.46	94.97	23.33	99.38
Grade 5 Spring	517	9	5	482	21	64.29	95.83	94.97	30.00	98.97

Table 36

FY11 Cross Validation Summary

R-CBM	*n*	True Positive	False Positive	True Negative	False Negative	% Positive Predictive Power	% Negative Predictive Power	% Overall Correct Classification	% Sensitivity	% Specificity
Grade 3 Fall	501	44	50	401	6	46.81	98.53	88.82	88.00	88.91
Grade 3 Winter	549	48	45	448	8	51.61	98.25	90.35	85.71	90.87
Grade 3 Spring	567	51	41	467	8	55.43	98.32	91.36	86.44	91.93
Grade 4 Fall	548	35	29	460	24	54.69	95.04	90.33	59.32	94.07
Grade 4 Winter	544	36	34	446	28	51.43	94.09	88.60	56.25	92.92
Grade 4 Spring	551	34	34	452	31	50.00	93.58	88.20	52.31	93.00
Grade 5 Fall	510	34	15	430	31	69.39	93.28	90.98	52.31	96.63
Grade 5 Winter	491	36	15	403	37	70.59	91.59	89.41	49.32	96.41
Grade 5 Spring	491	41	24	397	29	63.08	93.19	89.21	58.57	94.30

Chapter 5: Discussion of Findings, Conclusions, and Recommendations

Introduction

This final chapter includes a discussion of key findings, limitations, conclusions. Recommendations for policy and practice and for further study are also presented.

Study Purpose, Research Questions, and Design Overview

The purpose of this study was to explore the relationship between student performance on R-CBMs and student performance on the Alaska's SBA administered to students in Studied School District Grade 3 through Grade 5 students in the Studied School District as required by Alaska's accountability system. Two broad research questions framed this study across Grades 3, 4, and 5 and applied to interval data obtained through the triennial administration of R-CBM:

1. To what extent, if at all, is there a relationship between student performance on R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and student performance on the Alaska SBA administered in the spring of the same school year in the SSD?
2. To what extent, if at all, can cut scores be derived for each of the three R-CBM testing windows in the fall, winter, and spring that predict success on the Alaska SBA administered in the spring of the same school year in the SSD?

This study was non-experimental and correlational and was divided into four phases. The first phase relied on descriptive statistics to assist in determining the normality and distribution of data. Second, the relationship between student performance on R-CBM tools and student performance on the Alaska SBA was examined using Pearson correlation analysis. Third, binary logistic regression was used to determine cut scores for educators desiring to predict student outcomes on the Alaska SBA based on R-CBM performance. In the last phase, the cut scores established for testing data gathered during the fiscal year 2010 (FY10) school year were used to predict student success on the Alaska SBA administered during the fiscal year (FY11) school year.

Key Findings

The first research question sought to determine the extent of the relationship, if at all, between student performance on the R-CBM administered during the fall, winter, and spring in Grades 3, 4, and 5 and student performance on the Alaska SBA in the SSD. A statistically significant relationship between R-CBM and Alaska SBA was observed. The researcher determined efficacy for deriving cut scores via logistic regression for use in predicting whether students are on track to meet proficiency requirements on the Alaska SBA. Cross validation of all scores was necessary to determine the validity of using cut scores to make predictions on successive administrations of the Alaska SBA (Table 37).

Table 37

Summary of Pearson r Correlation Coefficients Between Assessments by Grade

		R-CBM		
SBA Reading SS	*n*	Fall	Winter	Spring
3rd Grade	472	.689**	.700**	.728**
4th Grade	435	.714**	.718**	.719**
5th Grade	517	.706**	.712**	.717**

Note. ** indicates significance as less than .01.

Eight key findings resulted from the analysis of the data regarding both research questions:

1. Strong correlations between R-CBM and the Alaska SBA existed between all grade

level administrations of the R-CBM and same grade administration of Alaska SBA

2. Within each grade, the strongest correlations were found in the Spring.
3. The use of logistic regression demonstrated that statistically derived cut scores could be used to classify student outcomes on the Alaska SBA correctly.
4. The logistic regression model favored specificity over sensitivity (i.e. values for specificity ranged from 92% to 96% while values for sensitivity ranged from 20.93% to 51.22%).
5. A cross validation of statistically derived cut scores established with data from the FY10 school year yielded similar results with regard to specificity as the logistic regression model displayed only a small accuracy decrease (≤ 2%).
6. A cross validation of statistically derived cut scores established with data from the FY10 school year yielded significantly different results. Specificity (89% to 97%) continued to be higher than sensitivity (49% to92%).
7. Overall correct classification was approximately 4% higher in the logistic regression model as compared to the classification of students in the cross validation.
8. Although statistically valid cut scores can be established, the cut score values were lower than cut scores established in different populations and exams other than the Alaska SBA.

Discussion of the Findings

Findings related to the first research question determined the presence of a relationship between student performance on R-CBM tools administered in Grades 3, 4, and 5 in the fall, winter, and spring and student performance on the Alaska SBA administered in the spring of the same school year in the SSD. Correlations ranged from

.689 - .728 with the strongest correlations found during the spring administration of the R-CBM. This might be attributed to students having increased reading rates overtime resulting from the additional instruction received throughout the school year thus increasing the likelihood passing the Alaska SBA as time spans between the administration of R-CBM and administration of the Alaska SBA decrease. These findings are consistent with correlations reported in similar studies. Shapiro et al. (2006) reported correlations between fifth grade CBM oral reading fluency and the Pennsylvania System of School Assessment (PSSA). With the exception of the fall assessment at one district, all correlations by Shapiro et al. ranged from .62 and .69. Third grade correlations between reading fluency and the Minnesota Comprehensive Assessment ranged from .68 in the fall to .71 in the spring (Silberglitt & Hintze, 2005).

Findings related to the second research question determined to what extent, if at all, the cut scores could be derived for each of the three R-CBM testing windows in the fall, winter, and spring to predict success on the Alaska SBA in the spring of the same school year in the SSD. CBMs have been widely used for predicting whether students are on track in reading. Much of the research has relied on establishing norms which have subsequently been generalized for use in varied populations. Currently, there is no universal definition by which students' progress can be determined however, NCLB mandates that states have integrated assessments that provide SEA and LEAS a construct for demonstrating growth against adopted standards. Researchers have begun looking to the CBM as a predictor variable when monitoring both district and student progress toward adequate yearly progress (AYP). The statistical model in this study was able to classify correctly the true positives, false positives, true negatives, and false negatives

93% to 96% of the time with regard to their associated test scores. Cut scores in the logistic model led to higher levels of specificity (94% - 100%) then sensitivity (21% - 51%) this is favorable in that the numbers of students who fail but were predicted to pass is minimized. Although statistical modeling had favorable results, it was necessary to cross validate the cut scores against the next year's data to determine if similar classifications could be expected if the scores were adopted. The overall correct classification (OCC) dropped by approximately 4% when the statistically derived cut scores were applied to the next year's student performance on the R-CBM and the ASBA. Although the OCC varied by only 4% in the cross validation, there were substantial differences in the levels of specificity (89% - 97%) and sensitivity (49% - 88%). The relatively high percentages of students that pass the ASBA, can account for the small changes in OCC as compared to more significant changes in sensitivity and specificity.

Conclusions

Two conclusions were drawn from the analysis of data and interpretation of findings. First, this study extended findings from previous studies by demonstrating that student performance on R-CBM is strongly correlated to student performance on state mandated assessments- specifically the Alaska Standards Based Assessment. This study provides further evidence that R-CBM can be used as a brief and efficient measure of student growth towards either established norms or specific targets set by individual school districts (Good et al., 2001; Deno, S., 2003; Deno et al., 2001; McGlinchey & Hixson, 2004; Shapiro et al., 2006; Shinn, 2007). Correlations for third graders (n = 472) ranged from .689 to .728; fourth grade correlations (n = 435) ranged from .714 to .719

and fifth grade correlations (n = 517) ranged from .706 to .717. Strong correlation between the two measures, validate findings of previous research and further identified the utility and value of R-CBM as a screening tool within an RTI construct (Ardoin & Christ, 2008; Crawford et al., 2001; Deno, S., 2003; Fuchs & Fuchs, 2004, 2007; Shinn, 2007; Silbert et al., 2006).

Second, this study demonstrated that statistically derived cut scores can be established which accurately measure progress towards the ASBA. This study established cut scores via logistic regression that predicted passing rates on the ASBA with 92% to 99% accuracy. NCLB requires each state to establish specific levels of proficiency. Silberglitt and Hintze (2005) suggested that cut scores can be established based on a single criterion, such as student outcomes on state mandated assessments, and this study demonstrated their suggestion's prudence. Further, this study demonstrated that a school district can use extant data to establish cut scores which in turn can be used to evaluate program effectiveness as well as resource allocation. Due to the ability of the cut scores established in this study to predict student performance accurately, the SSD may evaluate overall program and curriculum effectiveness based on the percentage of students predicted to be on track to pass the ASBA. Although cut scores were established that accurately predicted student progress for passing the ASBA, the scores were lower than scores found in previous studies (Crawford et al., 2001; Good et al., 2001; Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Sibley et al., 2007; Silberglitt & Hintze, 2005; Stage & Jacobsen, 2001). For this reason, practitioners should be cognizant that scores established in this study. Though accurate in predicting student outcomes on the

ASBA, the scores may not be rigorous enough to use for goal setting as it appears that the ASBA is less rigorous than other state exams.

Limitations Observed through Data Collection and Analysis

Even though the results did demonstrate that cut scores could be established to predict student success on the Alaska SBA, the scores were lower than scores observed in previous studies. The demographics for the students involved in this study were not representative of the demographic makeup of most of Alaska. One hundred percent participation for each grade did not occur for reasons beyond the control of the SSD; such lack of participation could have occurred for children suffering from illness, absenteeism, and transience. R-CBM, while representative of reading fluency, also addresses accuracy, and accuracy was not a variable addressed in this study. This study's findings might not be generalizable, and LEAs would be wise to determine cut scores locally.

Recommendations for Policy and Practice

The results of this study demonstrated that as in previous studies, R-CBM produces significant correlations to student performance on high-stakes exams. This attempt to establish cut scores for decision making within an RTI construct has potentially presented more questions than answers. Although statistically derived cut scores were established with great specificity (i.e., the ability to predict that a student will pass the Alaska SBA), the level of sensitivity (i.e., the ability to predict student failure on the Alaska SBA) was not nearly as accurate. Scores that favor specificity over sensitivity can be considered favorable to those that favor sensitivity in that most students who are predicted to pass have high odds of passing.

The inherent problem with the level of sensitivity occurs when allocating resources to the students predicted to fail the Alaska SBA. Using these predictions could lead to substantial numbers of students being labeled at risk even though the same students could likely pass the Alaska SBA without intervention. Although the current results suggested that cut scores could be used as a screening process, they were not as robust as scores determined in similar studies. For example, Good et al. (2001) established a score of 110 as a third grade benchmark in contrast to this study which established the cut score at 85 for the same period.

Because the scores in this study were significantly lower than those found in previous studies, they should not be viewed as a benchmark to which to strive but rather as a score with utility in dichotomizing a population of students as follows: Most students who score at or above the cut score will likely pass the Alaska SBA; most students who ultimately fail the Alaska SBA will likely be identified as students who fail to attain the cut score. One caution to this prediction is that there will likely be a significant number of students who are predicted to fail but do not, resulting in more students receiving remediation than necessary. The use of logistic regression yielded cut scores that favored specificity over sensitivity; however, providing more intervention to students who could also fail the Alaska SBA to ensure they will pass the assessment will ultimately benefit the students and the state.

Since increases or decreases in sensitivity and specificity are inversely related, costs occur when the goal is to maximize one over the other. When attempting to identify predicted failure, it is preferable to have higher levels of specificity over sensitivity. In other words, most students who are predicted to pass do so. Conversely, low levels of sensitivity can result in more students being predicted to fail than actually do. For this reason, additional measures may be needed to refine the screening process for identifying students at risk. These additional measures may require triangulating results with other data, using more complete prediction models, and strategic monitoring throughout the year.

R-CBM has consistently demonstrated usefulness as a screening tool and a measure of student performance. Current practice has extended the use of R-CBM to predict student outcomes on state exams. The use of a static score tied to a single outcome is desirable in that districts can establish cut scores specific to their populations and assessments. Although this study's findings demonstrated R-CBM as strongly correlated to the Alaska SBA and cut scores as useful for identifying as risk students, some caution is in order.

High percentages of students in the SSD continue to pass the Alaska SBA. The high pass rate coupled with cut scores from this study raise questions. Cut scores established using conditional probability, as is the case with logistic regression, are highly sensitive to the state pass rate. Daniel (2000) illustrated an inverse relationship between the percentage of students who passed a state test and the cut score needed to predict whether students would pass the state test. This finding might in part be reflected by the cut scores established in this study.

Although the cut scores established in this study demonstrated utility through the cross validation between the fiscal years 2010 and 2011 data and although cut scores are highly predictive of students passing the state test, the specific cut scores might have limited value for determining if students are making sufficient grade level progress as other same grade students. This situation was represented in previous studies which resulted in higher cut scores. There is evidence that the Alaska SBA lacks rigor, which might account for the unexpectedly low scores determined in this study. Given concerns regarding the rigor of the Alaska SBA as well as whether the cut scores established in the study are robust enough to use for goal setting, SSD should consider using alternative methods when establishing goal setting measures (e.g., the use of district of national grade level mean could serve as an appropriate alternative for use in goal setting).

Recommendations for Further Study

This study demonstrated strong positive correlations between R-CBM and Alaska SBA. Moreover, the study demonstrated that statistically derived cut scores could be used for predicting whether or not students are likely to pass the Alaska SBA. Based on these findings, the recommendations for future studies follow:

- Additional studies should be completed in Alaska with populations more reflective of the demographic found throughout the state to determine whether cut scores are consistent across changes in population demographics.
- All analyses in this study were conducted on the aggregate of all subgroups. Additional evaluation of R-CBM should be considered with specific emphasis on a strand analysis of ELL, SPED, and economically disadvantaged students, since these subgroups of students consistently underperform on the Alaska SBA.
- Further analysis or study should be completed in Alaskan school districts with lower pass rates on the Alaska SBA to validate further the finding by Daniel (2010) that scores based on conditional probability are highly sensitive to state-test pass rate.

- Logistic regression models favor specificity over sensitivity. Additional measures are needed to classify students that do not achieve established cut scores. Further study should be completed using complex logistic regression analysis using at least two predictor variables, the R-CBM and an additional locally relevant variable. A multivariate analysis may provide additional information and further refine the identification of students needing additional supports.
- A study to address how students predicted to fail versus pass are separated from each other and what RTI methods will prove more beneficial to these students is needed.
- Further study should be completed in states with similar pass rates on their respective state exams to determine why cut scores established in this study appeared to be lower than the cut scores reported in previous studies.

Chapter Summary

This study has built on previous research in other states by demonstrating that statistically derived cut scores can be established that accurately predict student performance on future assessments. This is especially useful within a RTI construct as it allows schools an efficient mechanism to identify students that may not be on track to pass the state exam. Though the overall classification of student is high, care must be given that regardless of the cut score, there will continue to be students predicted to pass which do not (false negatives). As a result, some students needing interventions may not be identified. For this reason, districts may need to consider secondary scores within a range and monitor students more frequently in an effort to prevent false negatives. Similarly, some students may predict to fail but go on to pass the Alaska Standards Based

Assessment (false positives). This may contribute to students receiving interventions unnecessarily.

With the ability to use cut scores which accurately identify student outcomes on the Alaska Standards Based Assessment, the SSD or other districts wishing to set cut scores should consider using a combination of both criteria and normative use of Reading Curriculum Based Measurements performance data. At a district level, if 80% of students do not routinely reach proficient on the Alaska Standards Based Assessment, then a program or curricular change may be in order. Normative data on the other hand can and should be used for resource allocation. While resources can be limited, using normative data to match available resources to the lowest performing students should be considered along with efforts to strengthen to core instruction in the regular classroom or tier 1.

While the findings of this study are specific to the SSD, specific lessons learned along with methodologies used may apply to other districts or states when attempting to align local assessments to state mandated high stakes assessments.

REFERENCES

AIMSweb. (2011a). *AIMSweb CBM tools meet scientific standards for use in frequent progress monitoring.* Retrieved from http://aimsweb.com/index.php?mact=News,cntnt01,detail,0&cntnt01articleid=62&cntnt01detailtemplate=Press-Room&cntnt01returnid=39

AIMSweb. (2011b). *AIMSweb systems: Reading curriculum based measurement (Reading-CBM).* Retrieved from http://www.rti4success.org/tools_charts/popups_progress/aimsweb_OralReading_area.php

Alaska Department of Education & Early Development. (2005). *Questions & answers regarding Alaska assessments*. Retrieved from http://www.eed.state.ak.us/tls/assessment/sba/Fall05/Q&ATestDevelopment.pdf

Alaska Department of Education & Early Development. (2007). *History of Alaska school reform 1991-2006.* Juneau, AK: Author. Retrieved from http://www.eed.state.ak.us/publications/historyreform.pdf

Alaska Department of Education & Early Development. (2011a). *Participation guidelines for Alaskan students in state assessments*. Retrieved from http://eed.alaska.gov/tls/assessment/pdf_files/ParticipationGuidelinesWeb_2011.pdf

Alaska Department of Education & Early Development (2011b). *Alaska Standards based assessments (SBA) operational and field test*. Retrieved from http://www.eed.state.ak.us/tls/assessment/sba.html.

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology*

Review, 37(1), 109-125. Retrieved from http://www.nasponline.org/publications/ spr/index.aspx?vol=37&issue=1

Austin, B., Mattos, M., & Weber, C. (2009). *Pyramid response to intervention: RTI, professional learning* communities*, and how to respond when kids don't learn.* Bloomington, IN: Solution Tree Press.

Barnett, D. W., Daly, E. J., III, Jones, K. M., & Lentz, F. E., Jr. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *Journal of Special Education, 38*(2), 66-79. Retrieved from http://www.ingentaconnect.com/content/proedcw/jse

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., . . . Tilly, W., D., III. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education. Retrieved from http://www.casecec.org/pdf/rti/RtI%20An%20 Administrator's%20Perspective%201-061.pdf

Bender, W. N., & Shores, C. (2007). *Response to intervention: A practical guide for every teacher*. Arlington, VA: Council for Exceptional Children.

Bender, W. (2009). *Beyond the RTI pyramid: Solutions for the first years of implementation.* Bloomington, IN: Solution Tree Press

Bergan, J. R. (1977). *Behavioral consultation.* Columbus, OH: Charles E. Merrill.

Berkley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 19*, 579-586. doi:10.1177/0022219408326214

Bradley, R., & Danielson, L. (2004). The office of special education program's LD initiative: A context for inquiry and consensus. *Learning Disability Quarterly, 27*(4), 186-188. Retrieved from http://www.questia.com/library/1G1-125647529/the-office-of-special-education-program-s-ld-initiative

Brigman, G., Webb, L., & Campbell, C. (2007). Building skills or school success: Improving academic and social competence. *Professional School Counseling, 10,* 279-288. Retrieved from http://www.mendeley.com/catalog/building-skills-school-success-improving-academic-social-competence-students-9/

Brown-Chidsey, R., & Steege, M. (2010). *Response to intervention: Principles and strategies for effective practice.* New York, NY: Guilford Press.

Buffum, A. G., Mattos, M., & Weber, C. (2009). *Pyramid response to intervention: RTI, professional learning communities, and how to respond when kids don't learn*. Bloomington, IN: Solution Tree Press.

Burns, M. K., & Gibbons, K. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. New York, NY: Routledge.

Burns, M. K., Tucker, J. A., Hauser, A., Thelen, R. L., Holmes, K. J., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: A potential criterion for curriculum-based assessments. *Assessment for Effective Intervention, 28*(1), 1-7. doi:10.1177/073724770202800101

Castillo, J. M., & Batsche, G. M. (2012). Scaling up Response to Intervention: The Influence of Policy and Research and the Role of Program Evaluation. *NASP*

Communiqué, *40*(8), 14-16. Retrieved from http://www.nasponline.org/publications/cq/40/8/scaling-up-response.aspx

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*(1), 130-146. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=36&issue=1

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*(4), 303-323. doi:10.1207/S15326977EA0704_04

Daniel, M. H. (2010). *Setting CBM benchmark targets for state-test success: Methodological challenges*. Paper presented at the annual meeting of the National Association of School Psychologists, Chicago, IL.

Deno, E. (1970). Special education as developmental capital. *Exceptional Children, 37*, 229-237. doi:10.1177/002246699402700402

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*(3), 184-192. doi:10.1177/00224669030370030801

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507. Retrieved from http://mdestream.mde.k12.ms.us/sped/ToolKit/Articles/Assessment/Deno.pdf

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children. Retrieved from http://www.cehd.umn.edu/EdPsych/rips/Documents/Data-Based%20 Program%20Modification-%20A%20Manual.pdf

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6*(4), 50-70. doi:10.1111/j.1937-8327.1993.tb00605.x

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Dynda, A. M. (2006). Integration of response to intervention and norm-referenced tests in learning disability identification: Learning from the tower of Babel. *Psychology in the Schools, 43*(7), 807-825. doi:10.1002/pits.20190

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=33&issue=2

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488-500. Retrieved from http://www.questia.com/library/1G1-10737382/paradigmatic-distinctions-between-instructionally

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61,* 15-24. Retrieved from http://www.questia.com/library/1G1-15819284/must-instructionally-useful-performance-assessment

Fuchs, L. S., & Fuchs, D. (2004). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. *Assessment for Effective Intervention, 29*(4), 25-37. Retrieved from http://www.ldonline.org/article/14601/

Fuchs, L. S., & Fuchs, D. (2006a). A framework for building capacity for responsiveness to intervention. *School Psychology Review, 35*(4), 621-626. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=35&issue=4

Fuchs, D., & Fuchs, L. S. (2006b). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99. doi:10.1598/RRQ.41.1.4

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 14-20. Retrieved from http://www.sde.idaho.gov/site/rti/resourcesDocs/General%20Resources/Teaching%20Exceptional%20Children.pdf

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

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

Fuchs, L. S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum based reading assessment. *Assessment for Effective Intervention*, *9,* 191-207. doi:10.1177/073724778400900401

Gallagher, D. (2010). Hiding in plain sight: The nature and role of theory in learning disability labeling. *Disability Studies Quarterly,30*(2). Retrieved from: http://dsq-sds.org/article/view/1231/1278

Glover, T. A., & DiPerna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review, 36*(4), 526-540. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?issue=4&vol=36

Good, R. H., III, Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257-288. doi:10.1207/S1532799XSSR0503_4

Griffiths, A., Parson, L. B., Burns, M. K., VanDerHeyden, A., & Tilly, W. D. (2007). *Response to intervention: Research for practice.* National Association of State Directors of Special Education. Retrieved from www.nasdse.org/Portals/0/Documents/RtI_Bibliography2.pdf

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in Grades 2 through 5. *Teaching Exceptional Children, 24*(3), 41-44.

Heller, K. A., Holtzman, W. H., & Messick, S. (1982). *Placing children in special education: A strategy for equity.* Washington, DC: National Academy Press.

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools*, *43*(1), 45-56. doi:10.1002/pits.20128

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34*(3), 372-386. Retrieved from http://www.nasponline.org/publications/spr/abstract.aspx?ID=1761

Hosmer, D.W & Lemeshow,S. (2000). *Applied Logistic Regression* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Hunley, S. A., McNamara, K., & National Association of School Psychologists. (2010). *Tier 3 of the RTI model: Problem solving through a case study approach*. Thousand Oaks, CA: Corwin Press.

Individuals with Disability Education Act Amendments of 1997. (1997). Retrieved from http://thomas.loc.gov/home/thomas.php

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 105-17, 118 Stat. 2647 (2004).

Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). *Responsiveness to intervention (RTI): How to do it*. Lawrence, KS: National Research Center on Learning Disabilities. Retrieved from http://www.nrcld.org/rti_manual/

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*(4), 174-185. doi:10.1111/j.1540-5826.2009.00291.x

Kavale, K. A., & Spaulding, L. S. (2008). Is Response-to-Intervention good policy for specific learning disability? *Learning Disabilities Research & Practice, 23*(4), 169-179. doi:10.1111/j.1540-5826.2008.00274.x

Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future.* Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Education_Topics/pdf/AmericasPerfectStorm.pdf

Kovaleski, J. F. (2007). Response to intervention: Considerations for research and systems change. *School Psychology Review, 36*(4), 638-646. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?issue=4&vol=36

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Marston, D. B. (1989). Curriculum-based measurement: What it is and why we do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessment special children* (pp. 18-78). New York, NY: Guilford Press.

McBeath, J., Reyes, M. E., & Ehrlander, M. (2008). *Education reform in the American states*. Charlotte, NC: Information Age. Retrieved from http://books.google.com/books?id=Qtn3gNwqk6wC&printsec=frontcover#v=onepage&q&f=false

McCook, J. E. (2006). *The RTI guide: Developing and implementing a model in your schools*. Horsham, PA: LRP.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*(2), 193-203. Retrieved from http://www.wce.wwu.edu/Depts/SPED/Forms/Kens%20Readings/CBM/Readings/CBM%20Using%20CBM%20to%20predict%20performance%20on%20state%20assessments%20in%20reading%20McGlinchey%202004.pdf

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education, 41*(2), 68-84. Retrieved from http://www.eric.ed.gov/PDFS/EJ775111.pdf

Minitab. (2010). *Creating a receive operating characteristic (ROC) curve from a binary logistic regression model.* State College, PA: Minitab. Retrieved from http://www.minitab.com/en-US/support/answers/answer.aspx?ID=2513&langType=1033

National Association of State Directors of Special Education. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: Author.

National Association of State Directors of Special Education. (2008a). *Response to intervention: Blueprints for implementation.* Alexandria, VA: Author.

National Association of State Directors of Special Education. (2008b). *Response to intervention: Policy consideration and implementation.* Alexandria, VA: Author.

National Association of State Directors of Special Education, & Council of Administrators of Special Education. (2006). *Response to intervention: NASDSE and CASE white paper on RTI.* Alexandria, VA: Author. Retrieved from http://www.nasdse.org/Portals/0/Documents/Download%20Publications/RTIAnAdministratorsPerspective1-06.pdf

National Center for Education Statistics. (2003). *The nation's report card: Reading highlights 2003*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2003/2004452.pdf

National Center on Response to Intervention. (2010). *Essential components of RTI–A closer look at response to intervention.* Retrieved from http://www.rti4success.org/pdf/rtiessentialcomponents_042710.pdf

National Center on Student Progress Monitoring. (n.d.). *Homepage*. Retrieved from http://www.studentprogress.org/

National Joint Committee on Learning Disabilities. (2005). *Responsiveness to intervention and learning disabilities.* Retrieved from http://www.ldonline.org/article/11498/

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED444126&ERICExtSearch_SearchType_0=no&accno=ED444126

No Child Left Behind Act (NCLB) of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Ofiesh, N. (2006). Response to intervention and the identification of specific learning disabilities: Why we need comprehensive evaluations as part of the process. *Psychology in the Schools, 43*(8), 883-888. doi:10.1002/pits.20195

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families*. Retrieved from www2.ed.gov/inits/commissionsboards/whspecialeducation/reports/summ.html

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests. *Journal of Psychoeducational Assessment, 24*(1), 19-35. doi:10.1177/0734282905285237

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*(4), 601-617. Retrieved from http://www.nasponline.org/publications/spr/abstract.aspx?ID=1834

Shinn, M. R., Good, R. H., III, Knutson, N., Tilly, W. D., & Collins, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory factor analysis of its relation to reading. *School Psychology Review, 21*, 459-479. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=21&issue=3

Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb® training workbook: Administration and scoring of reading maze for use in general outcome measurement*. Retrieved from http://www.aimsweb.com/uploads/pdfs/scoring_maze.pdf

Shores, C., & Chester, K. (2009). *Using RTI for school improvement: Raising every student's achievement scores*. Thousand Oaks, CA: Corwin Press.

Sibley, D., Biwer, D., & Hesch, A. (2001, April). *Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement.* Paper presented at the annual conference of the National Association of School Psychologists, Washington, DC.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*(5), 527-535. doi:10.1002/pits.20175

Silberglitt, B., & Hintze, J. (2005). Formative assessment using R-CBM cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23*(4), 304-325. doi:10.1177/073428290502300402

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*(3), 407-419. Retrieved from http://www.nasponline.org/publications/spr/abstract.aspx?ID=1573

Stanovich, K. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. *Remedial and Special Education, 5*(3), 11-19. doi:10.1177/074193258400500306

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure, 52*(2), 48-58. doi:10.3200/PSFL.52.2.48-58

Swets, J. A. (1998). Measuring the accuracy of diagnosis systems. *Science, 240,* 1285-1293.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1-26. doi:10.1111/1529-1006.001

Tilly, W. D. (2003, December). *How many tiers are needed for successful prevention and early intervention? Heartland Area Education Agency's evolution from four to three tiers.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.

Tindal, G., Marston, D., & Deno, S. (1983). *The reliability of direct and repeated measurement* (Research Report No. 109). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.

United States Commission on Civil Rights. (2002). *Racism's frontier: The untold story of discrimination and division in Alaska*. Washington, DC: Author.

University of Oregon. (n.d.). *DIBELS oral reading fluency retell fluency.* Retrieved from http://dibels.uoregon.edu/measures/orf.php

U.S. Department of Education. (2002). *No child left behind: A desktop reference*. Retrieved from www2.ed.gov/admins/lead/account/nclbreference/ reference.pdf

VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children*, *77*(3), 335-350. Retrieved from http://journals.cec.sped. org/ec/all_issues.html

VanDerHeyden, A. M., Witt, J. C., & Barnett, D. W. (2005). The emergence and possible futures of response to intervention. *Journal of Psychoeducational Assessment, 23*(4), 339-361. Retrieved from http://abbyowens.wiki.westga.edu/file/view/ technical+adequacy+of+rti+decisions.pdf

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. Retrieved from http://www.disfor.unict.it/Public/Uploads/links/redefining%20learning%20dis.pdf

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*(2), 85-104. doi:10.1207/s15326977ea1102_1

Yeo, S. (2010). 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

APPENDIX A

Grade 3 R-CBM Text

"Where are you going, Dad?" I ask excitedly. I wonder if something 12
interesting is happening. 15

"I'm going to search for some deer. Would you like to come along? 28
We'll take a trek in the woods," replies Dad. 37

"I love going for walks. Wait for me!" I reply. 47

"I want to go too!" yells Mike, my younger brother. "Please help me 60
tie my shoes!" 63

"Don't worry, Mike. I will help you. Dad always waits for both of us," 77
I explain calmly. 80

We live in the country with huge trees behind our house. During the 93
different seasons of the year, my brother and I like to walk along the 107
paths that go through the trees. Dad usually goes with us and teaches 120
us things about nature. 124

It's a fall afternoon and our shuffling feet make quite a racket 136
through the dry leaves. Dad tells us to try to be quiet. He doesn't want 151
us to scare the deer away. 157

"Shhhh!" says Dad. "Stop and listen!" 163

My little brother and I stop, but we don't hear anything. 174

"I hear something!" whispers Mike. "Over there!" he points. 183

I look to where he's pointing and see a big, brown deer looking right 197
at us! She isn't moving, but her head is up high. She's listening just like 212
we are! The deer puts her head down, grunts, and stomps her front 225
hoofs on the ground. We wait while Dad smiles and lifts his camera to 239
his face. Click! ... whirr ... Click! Dad takes two pictures. 248

Two smaller deer stand behind the doe! They are her baby fawns, 260
born last spring. They are eating acorns off the ground. The fawns 272
don't even see us! The doe snorts again and turns to jump away. The 286
two little deer follow her. 291

"That was really cool, Dad. Thanks for taking us with you," we say. 304

APPENDIX B

Grade 3 Reading Assessment Alaska SBA

READING—SAMPLE QUESTIONS

Directions

Read the Sample Passage and Sample Questions A and B. For Sample A, fill in the circle that goes with the answer you choose. Be sure to fill in the circle completely and make your mark heavy and dark. If you want to change your answer, completely erase the mark you made before making a new mark. For Sample B, write your answer on the lines provided.

SAMPLE PASSAGE

The sun was shining brightly during recess. Anne, Dustin, Ivan, and Katia were playing soccer. Suddenly the school bell began to ring. They looked at each other in surprise. Recess couldn't be over already! The first kickoff had just taken place.

Ivan said, "There must be something wrong with the school bell. Why is it ringing now?"

"Let's go find out," said Anne.

So they all jogged toward the school doors leading to the office. The doors flew open just before they arrived. The custodian came running out holding something at arm's length in a net. It was wiggling and twisting this way and that. It was trying to escape. All that the students could see was a small ball of fur in the net.

Sample A

Why were the students first surprised?

- (A) The soccer game was canceled.
- (B) The school bell rang too early.
- (C) The custodian had something in a net.
- (D) The teacher gave them an extra recess.

GO ON

READING

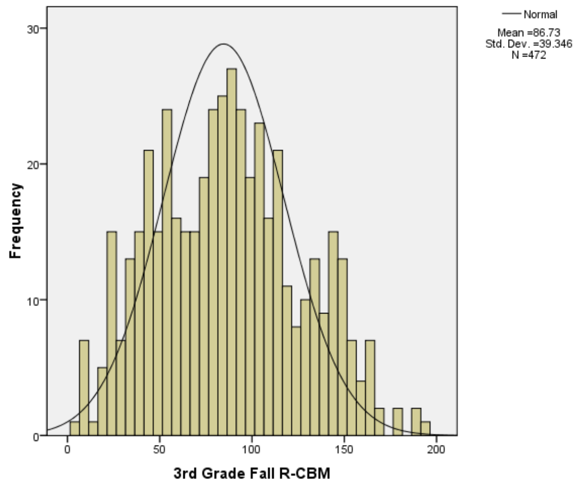
Sample B

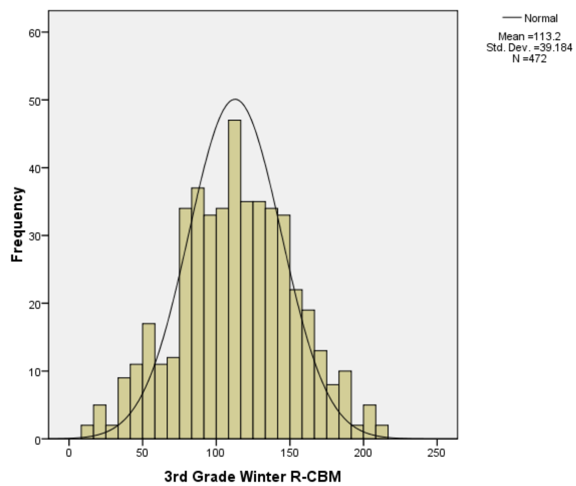
Tell what the custodian might have in the net. Use an example from the passage to support your response.

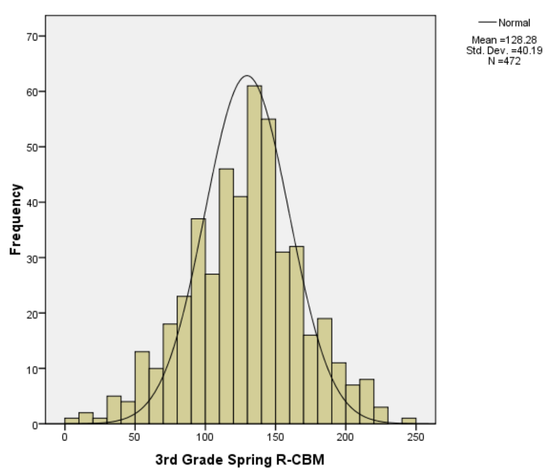
APPENDIX C

Histograms for R-CBMs by Grade and Time of Year

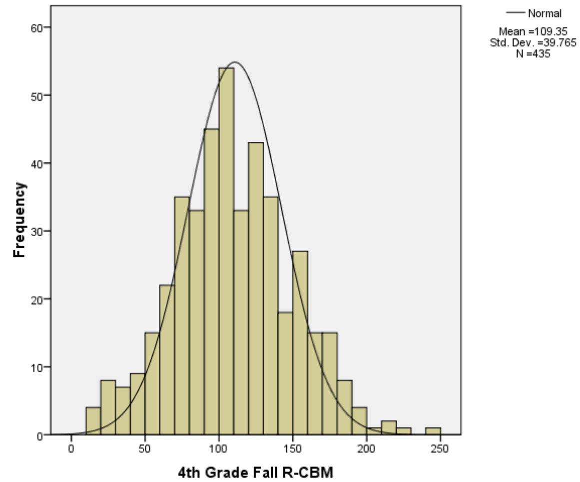
Grade 3 Fall, Winter, and Spring R-CBM Histograms

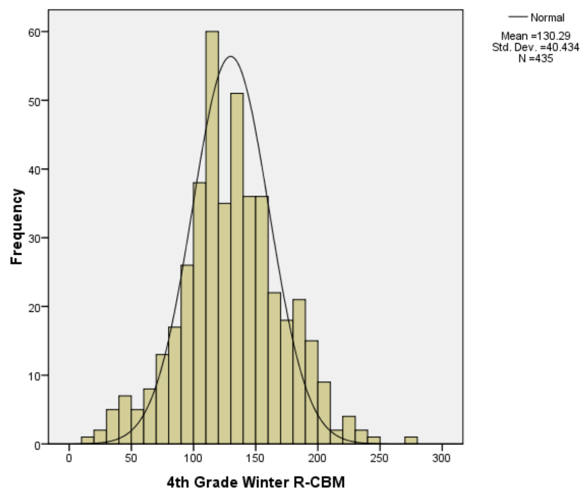


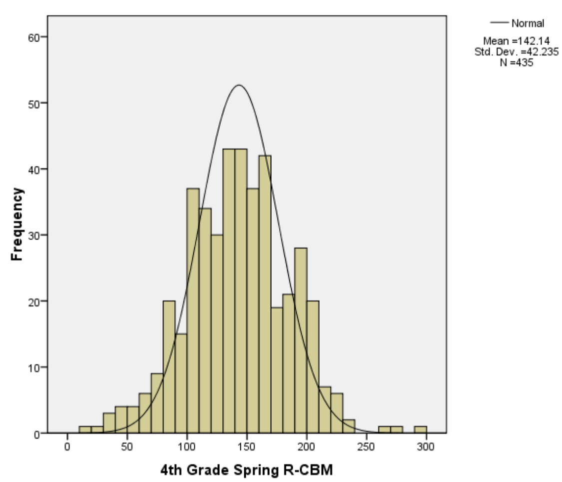




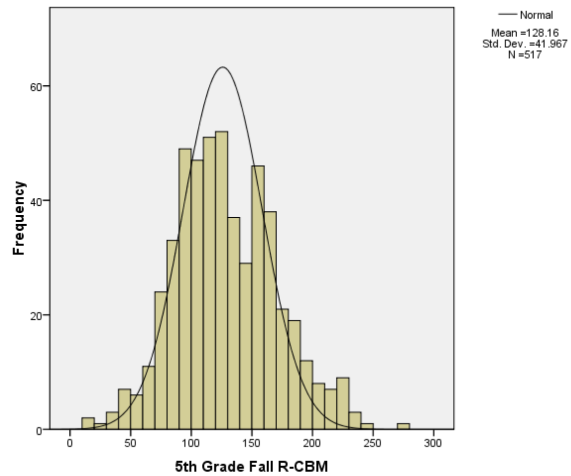
Grade 4 Fall, Winter, and Spring R-CBM Histograms

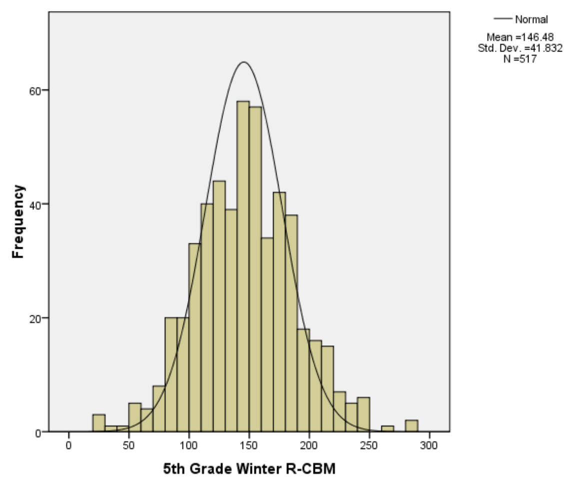


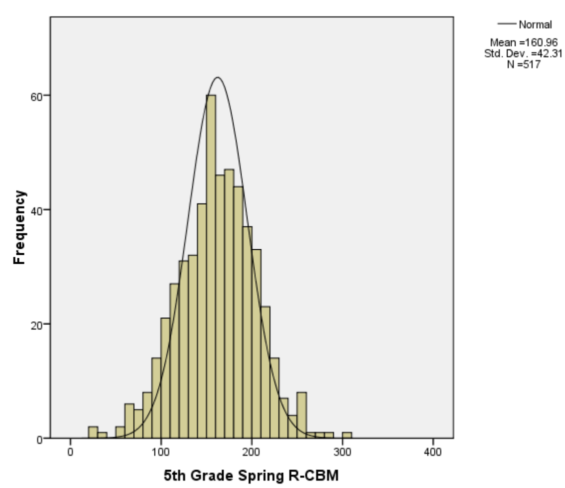




Grade 5 Fall, Winter, and Spring R-CBM Histograms







Histograms for Grades 3, 4, and 5 Spring ASBA Reading Scores