THE IMPACT OF THE ACCOUNTABILITY MOVEMENT ON PRINCIPAL EVALUATION: UNDERSTANDING THE ROLE OF FORMATIVE VERSUS SUMMATIVE ASSESSMENT

by

DAWN ELIZABETH STRONG

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Education
June 2016

DISSERTATION APPROVAL PAGE

Student: Dawn Elizabeth Strong

Title: The Impact of the Accountability Movement on Principal Evaluation: Understanding the Role of Formative Versus Summative Assessment

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

Charles R. Martinez	Chairperson
Keith Hollenbeck	Core Member
Joanna Smith	Core Member
Chris Murray	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

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

Degree awarded June 2016

DISSERTATION ABSTRACT

Dawn Elizabeth Strong

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

June 2016

Title: The Impact of the Accountability Movement on Principal Evaluation: Understanding the Role of Formative Versus Summative Assessment

This study analyzed the required inclusion of school test scores in the yearly evaluation of school principals within the current school reform and standards and accountability movements of both the federal and state departments of education. Extant data from a single school district in Oregon was used for this study, and included: (a) district-wide elementary principal summative performance evaluation scores, (b) district-wide fourth and fifth grade fall and spring reading curriculum-based measures scores curriculum-based measures, (c) 2013 and 2014 spring reading scores from the Oregon Assessment of Knowledge and Skills (OAKS-R), and (d) student demographic variables. The student non-academic predictor variables (demographic risk factors) included in the study were (a) attendance, (b) English Language Proficiency (ELP), (c) Free and Reduced Meals (FARMS), (d) percent Other-than-White, and (e) Special Education. Multiple regression analyses were used to determine which assessment and/or non-assessment factors accounted for differences between principals' summative evaluation scores. The results indicated that Summative Principal Ratings are poor predictors of the academic success of all students whether using large-scale summative assessment (OAKS-R) or formative assessments (easyCBM), with all measures only accounting for a

miniscule portion of the Summative Principal Rating variance. However, demographic variables were slightly more related to the Summative Principal Rating. Practical implications of using student test scores to hold principals accountable for the academic results of all students are discussed in relation to district administrative policy and placement procedures for administrators and teachers, examining the behaviors and practices of teachers' whose students have shown the most gains, and using these successful teacher practices a basis for teacher-to-teacher district-wide professional development. Finally, suggestions for future research in the areas of improving principal evaluation systems and the study of direct and indirect impacts principals have on student success and achievements are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Dawn Elizabeth Strong

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Oregon State University, Corvallis, Oregon
Southern Oregon University, Ashland, Oregon

DEGREES AWARDED:

Doctor of Education, Educational Methodology, Policy, and Leadership, 2016, University of Oregon
Master of Science, Counseling, 1997, Oregon State University
Bachelor of Science, Education, 1983, Southern Oregon University

AREAS OF SPECIAL INTEREST:

Educational Leadership
Human Resources in Education
Collective Bargaining and Negotiations

PROFESSIONAL EXPERIENCE:

Executive Director, Oregon School Personnel Association, 2015-current

Interim Director of Human Resources, Multnomah Education Service District, Portland, Oregon, 2015-current

Director of Human Resources, Springfield Public Schools, Springfield, Oregon, 2008-2015

Principal, Agnes Stewart Middle School, Springfield School District, Springfield, Oregon, 2004-2008

GRANTS, AWARDS, AND HONORS:

Teacher of the Year, Mapleton School District, 1993

A Champion in Education, Administrator of the Year, Springfield 2008

Human Resource Professional Community Service Award 2014

ACKNOWLEDGMENTS

I wish to express my genuine appreciation to Dr. Keith Hollenbeck, who has far exceeded my expectations as a friend, an advisor and mentor. Never did I imagine in 1998 when we began working together as colleagues in Springfield School District that 18 years later I would be thanking him for his ongoing support throughout my doctoral program. I will forever be grateful for his many hours of reading and editing my drafts, giving technical advice and suggestions, and most importantly his encouragement and belief in me. His patience, support, friendship and guidance have been immeasurable. I would like to thank my committee who helped me think critically about my research. I would like to thank Dr. Nancy Heapes for her quiet, guide-by-the-side support and friendship, and for challenging me personally and professionally. I would like to thank my friend Brooke Wagner, who invited me on this doctoral journey with her, for her friendship, support, and the never-ending craziness that we experienced in life and this program together. I would like to thank my friend Julie for her strength, steadiness, and support to our children and me when I worked late into the nights, on weekends, and days at a time. I would like to thank my friend Joy for her unwavering commitment and inspiration, to keep me going, and to see me through to the end. Finally, to my children Tyler and Kelsey, who when I asked them what they thought about me going back to school for my doctorate and continuing to work full-time while doing it, told me, “Go for it! If that’s what you want to do, then we support you.” You inspire me every day to be the very best I can be. I love being your mom, and I love you more than words could ever say.

For my mom who has always been there for me, and for my children who make me want to be a better person every day – you light up my world!

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Oregon's effort on school reform and the standards and accountability movements mirrors the federal government's efforts. The ESEA Flexibility Document, released by the US Department of Education in 2011 (U.S. Department of Education, 2012) offered each State educational agency (SEA), on behalf of the State and local educational agencies (LEA), the opportunity to request flexibility regarding certain aspects of the No Child Left Behind Act of 2001. States choosing to participate were to develop rigorous and comprehensive plans designed to improve educational outcomes for all students, close achievement gaps, increase equity, and improve the quality of instruction (U.S. Department of Education, 2013). The waiver would ensure that SEAs, LEAs and schools would hold all students to higher standards, while ensuring that low-achieving students in each subgroup were identified and receiving targeted interventions based on need. The waiver also challenged SEAs to have an effective principal in every school, and an effective teacher in every classroom. Supporting effective instruction and leadership, which required states to implement teacher and principal evaluation and support systems that provided feedback and support needed to improve practices and increase student achievement, was one of the waiver's four core principles (U.S. Department of Education, 2012). To meet the educator effectiveness requirements of the ESEA Flexibility Waiver, Oregon passed Senate Bill 290(Or. 2011), which called for Oregon to strengthen educator evaluations by adopting a statewide framework for educator effectiveness, that included core performance standards of evaluation for teachers and

educational leadership standards for administrators, and a new focus on student learning and growth (Hungerford & Dickson, 2012).

A Historical Context for SB 290

Understanding Oregon's Senate Bill 290 requires historical context. As part of his *War on Poverty*, President Lyndon B. Johnson signed the Elementary and Secondary Education Act (ESEA) in 1965; it still functions as the federal government's primary education law. "Since its inception, ESEA has consistently remained the single largest fiscal source of federal support for educationally vulnerable school children" (Thomas & Brady, 2005, p. 51). Over the past 50 years, presidential mandates and congressional reauthorizations have amended ESEA with the stated pursuit of improving academic outcomes for all students, decreasing the achievement gap, and increasing educator effectiveness and accountability.

Figure 1 summarizes various US presidential administrations' attempts to hold states, schools and educators accountable for the academic results of all students. While some may question the actual outcomes versus the intended outcomes of ESEA, recent federally funded research suggested that while the original goal of reducing the achievement gap has not been achieved, had it not been for Title I, the nation's underserved and at-risk students would have fallen further behind (McDonnell, 2005).

The next section provides a brief overview of presidential accountability systems across presidential administrations. I specifically review the progression of legislation beginning with the passage of the ESEA in 1965, to national concerns regarding student education and their impacts on legislation, to the reauthorizations of ESEA that occurred during the Clinton and G. W. Bush administrations, and the most recent ESEA legislation

of the Obama administration, the ESEA Flexibility Waiver.

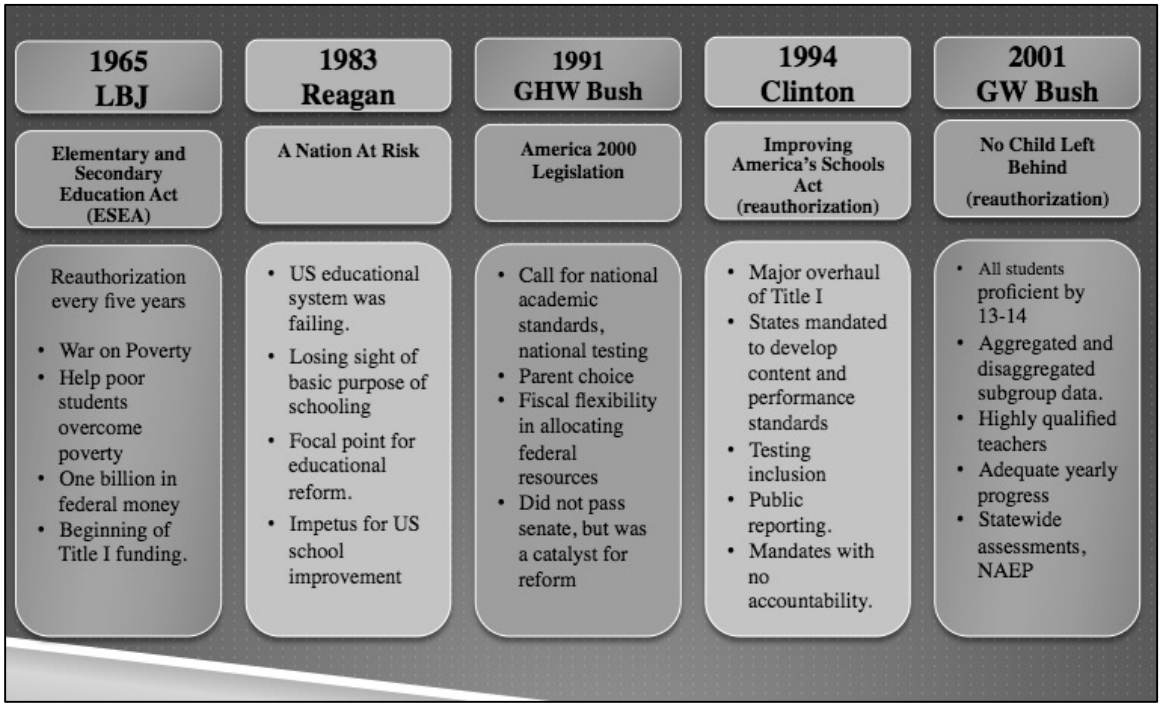

Figure 1. U.S. presidential administrations' most important attempts to hold states, schools, and educators accountable for the academic results of all students.

Lyndon B. Johnson: The Elementary and Secondary Education Act of 1965

On April 9, 1965 Congress enacted the Elementary and Secondary Education Act of 1965 (ESEA) (P.L. 89-10). ESEA brought attention to the educational needs of poor children, established federal standards for disadvantaged students, and symbolized the changing role of the federal government in education (Kanter, 1991). Through a special source of funding, known as Title I, Part A (Title I), the law allocated large resources to meet the needs of educationally deprived children, especially through compensatory programs for the children and families in poverty. Schools that received funds under Title I were required to be comparable in services to schools that did not receive Title I funds. The public policy purpose was to ensure federal financial aid was spent on top of state

and local funds, to which all public school children were entitled (McClure, 2008).

Congress reauthorized the Elementary and Secondary Education Act every five years through 1980, and yet this was not enough to stop a general concern that occurred in the late 1970s and early 1980s. There was widespread perception that the U.S. educational system was falling short of the implicit goal of keeping American students better educated than students in the rest of the world (Gardner, Larsen & Baker, 2005). Longtime U.S. industries were becoming challenged by high-quality products produced less expensively overseas; many believed this was due to American students falling behind their foreign counterparts in learning the skills necessary to keep the American economy afloat. Consequently, the federal government initiated steps to examine the quality of the education students in U.S. schools were receiving. On August 26, 1981, then Secretary of Education T. H. Bell created the National Commission on Excellence in Education to investigate the quality of education in the United States and to make a report within 18 months of the first meeting. Research was commissioned and public hearings were held to gather information for the report (Gardner et al., 2005).

Ronald Reagan: A Nation at Risk

The genesis for this effort was based on the National Commission on Excellence in Education's (Commission, 1983) report that stated America was at risk of being overtaken by its global competitors. It inferred that America and its educational institutions had lost sight of the basic purposes of schooling and of the high expectations and disciplined effort needed to attain them. Indicators of student performance and the demand for new skills led some researchers to conclude that the United States was raising a scientifically and technologically illiterate generation. The Commission warned the

United States faced a grave risk of losing its leading position in the world. “We had little idea of how we were doing, and we were happily complacent in assuming that we had, and would continue to have, the best schools money could buy” (p. 112). *A Nation at Risk* (U.S. Department of Education, 2008) highlighted a variety of challenges, findings and recommendations. About 13% of 17-year-olds were functionally illiterate, Scholastic Aptitude Test (SAT) scores were dropping, and the number of students who needed remedial courses in college was increasing. Such trends were viewed to be threatening children’s opportunities and the collective future of the United States. Recommendations for schools included (a) strengthening and increasing high school curriculum and graduation requirements, including courses in English, math, social studies, computer science, and foreign language, (b) devoting considerably more time to learning the *New Basics*, which would require a more-effective use of the existing school day, a longer school day, or a lengthened school year, and (c) a seven-part recommendation for teaching that was intended to improve the preparation of teachers, and to make teaching a more respected profession. The recommendations for colleges and universities were to raise admission requirements and expectations for academic performance and student conduct, and to adopt more rigorous and measurable standards. Two final recommendations of the Commission (1983) were for citizens across the nation to hold educators and elected officials responsible for providing the leadership necessary to achieve reform, and to provide the fiscal support and stability required to bring about the proposed reforms. *A Nation at Risk* became a focal point for state and local education agencies to undertake reform, while the government’s role was less prominent. According to Thomas and Brady (2005) the report had unquestionable policy

significance, and by the mid 1980's, 29 states required teachers to pass mandatory, standardized tests to gain certification, and 41 states had adopted increased academic requirements for high school graduation.

George H. W. Bush: America 2000

President Bush unsuccessfully attempted to advance a major legislation initiative known as America 2000, which called for national academic standards and national testing of students. According to McDonnell (2005) the failure was the result of Republicans who were opposed to any increase in the federal government's role in public education. However, the failed passage was still a catalyst for education reform and the development of academic standards common to all students.

William J. Clinton: Improving America's Schools Act

The major education reform initiative of the Clinton presidency was the Goals 2000: Educate America Act (Public Law 103-227, 1994), which continued with the standards-based education reforms of the previous administration. It included a focus on student achievement, challenging academic standards that would include all students, and a reliance on achievement testing as a means to monitor reform effects (McDonnell, McLaughlin, & Morison, 1997). In 1994 the passage of the Improving America's Schools Act (Public Law 103-328, 1994) led to the reauthorization of ESEA, whose purpose was to enable schools to provide opportunities for all children served to acquire the knowledge and skills contained in challenging State content standards and to meet State performance standards. Districts were required to identify schools not making *adequate yearly progress* (AYP*)* and formal steps were to be taken to improve them. McDonnell (2005) reported that as a precondition of receiving Title 1 funds, states were mandated to

demonstrate that learning goals, academic expectations, and curricular opportunities were the same for all students, regardless of whether or not they were eligible for Title 1 funds.

George W. Bush: No Child Left Behind

The last reauthorization of ESEA occurred in 2002, when Congress amended and reauthorized the Elementary and Secondary Act, and President George Bush signed into law the No Child Left Behind Act of 2001 (No Child Left Behind [NCLB], 2002). Beimers (2008) identified the four pillars of NCLB as (a) stronger accountability for results, (b) more freedom for states and communities, (c) proven education methods, and (d) more choices for parents. Borkowski and Sneed (2006) believed that NCLB's most important potential benefit was not only the fundamental belief that all children could learn, but that all children had the right to be taught. However, they also believed that unless fundamental changes were to occur, NCLB's ability to influence reform would be inconsistent at best. According to McDonnell (2005), the passage of NCLB was seen as a way to deal with a federal education policy that had provided billions of dollars in support, but had not demanded real results. McDonnell cited a synthesis of 17 federally commissioned Title I studies, with results that suggested while the original goal of closing the achievement gap had not been achieved, had it not been for Title I, the nation's underserved and at-risk students would have fallen further behind academically. The findings of these empirical studies also indicated "although Title I has produced a modest effect on students' yearly achievement gains, the effect has been extremely variable across subject areas, testing cycles, grade levels, and schools" (McDonnell, 2005, p. 33). Instead of the elusiveness of Title I's effects, NCLB's public reporting requirement of aggregated and disaggregated data was identified as critical to ensuring

that all students, regardless of their background characteristics or special needs, receive the education for which they are entitled (Beimers, 2008).

Highly qualified teachers. Provisions of NCLB (2002) required newly hired teachers in Title I programs to be highly qualified prior to placement in the classroom, and a highly qualified teacher in every classroom by the end of the 2005-2006 school year. According to NCLB (2002) highly qualified teachers must meet three criteria: (a) have a bachelor's degree, (b) be fully certified or licensed, including certification obtained through alternative routes, but excluding licensure that has been waived on an emergency, temporary, or provisional basis, and (c) have demonstrated content knowledge in the subject he or she teaches (Section 9191; NCLB, 2001). While the federal government established the highly qualified standards for all newly hired Title I teachers, states were given flexibility to determine highly qualified standards for veteran secondary teachers. Teachers new to the profession were required to demonstrate a high level of competency by: (a) passing rigorous State academic subject area tests for each subject in which they wish to teach, (b) having at least an academic major in the subject they wish to teach, or (c) passing a state-determined test of subject matter content (Smith, Desimone & Ueno, 2005). Requirements to meet the content knowledge for veteran teachers required: (a) passing a state-determined test, (b) a college major in the subject area, or (c) meeting criteria through the High Objective Uniform State Standard of Evaluation (HOUSSE) provision. HOUSSE provided the opportunity for veteran teachers to demonstrate content knowledge through a combination of college coursework, professional development, and other state-determined measures (Smith et al., 2005).

Adequate yearly progress. NCLB (2002) required states to make a number of changes in their testing and accountability systems. There were no national standards; instead states were required to develop Adequate Yearly Progress (AYP) objectives that met the requirements of the law. Annual tests were to: (a) measure the achievement of all students; (b) align with the state's academic content and achievement standards; (c) be used only for purposes where the tests were valid and reliable; (d) be used only if the state's Department of Education provided evidence that the tests were of adequate technical quality; (e) be administered to students in grades 3-8, and one high school grade chosen by the state; and (f) include multiple up-to-date measures that assessed higher order thinking skills and understanding (Kucerik, 2002).

Annual reports indicating school performance and statewide progress were required, with test results broken down by demographic subgroup. By the end of the 2013-2014 school year, every child was to be proficient in reading and math, as measured by their state's accountability system. Ninety-five percent of students in each subgroup were required to participate in state assessments, and subgroup results were to be reported separately to identify whether or not each group met objectives. Schools that met or exceeded AYP objectives or closed achievement gaps were eligible for State Academic Achievement Awards, while schools that failed to meet AYP objectives for two consecutive years were identified for improvement (NCLB 2002).

NCLB (2002) relied on assessment and accountability as a mechanism for improving student achievement, but it also included penalties for schools that failed to meet the AYP objectives (McDonnell, 2005). Linn, Baker, and Betebenner (2002) agreed that holding schools accountable would contribute to improvement, but concluded that

the goal of 100% of students reaching the proficient level or higher was completely out of reach. They suggested that unobtainable goals would do more harm than good, and would result in demoralizing teachers instead of motivating them to a better job. “Goals need to provide a challenge but not be set so high that they are unachievable” (p. 12).

Accountability at the individual school level. Linn et al. (2002) suggested school-level AYP objectives presented several substantial challenges. The variability differences in cohorts of students, changes in teaching staff, and the reliability of the assessments, measurement issues, and sampling errors could impact schools’ abilities to meet AYP, which would lead to a series of consequences that become more stringent over time. NCLB (2002) expected that schools meet AYP objectives, and at the end of two years, school-level results would be available in order to identify schools in need of improvement. Dee and Jacob (2010) explained that states were required to introduce sanctions and rewards based on every school’s AYP status. Explicit mandates included increasingly severe sanctions for persistently low-performing schools receiving Title I funding, from implementing school choice, to staff replacement and school restructuring.

Barack Obama: Race to the Top

On January 20, 2009 Barack Obama was inaugurated as the 44th President of the United States. One of his first initiatives was a $787 billion stimulus package known as the American Recovery and Reinvestment Act (ARRA [Pub. L. 111-5] 2009), which provided an unprecedented $100 billion in federal aid for education through the State Fiscal Stabilization fund and the Race to the Top (RTTT) fund.

Race to the Top. The $4.35 billion RTTT funds were the largest, federal education grant funds that an education secretary ever had the power to distribute. Funds

were not allocated, but distributed as the result of a highly competitive grant application process, and were intended to spur innovation and reform in state education policy and to create a new set of assessments aligned to new common academic standards (Whilden, 2010). On November 19, 2009 the RTTT rules were published. States who wished to compete for the money had to show how they would support (a) efforts in the development of internationally benchmarked academic standards and student assessments; (b) teacher and principal recruitment, development and retention; (c) construction of state systems that would link data on student success on statewide assessments to information about teachers and school practices across time in a system referred to as *value-added*; and (d) efforts to turn around the lowest performing schools. States were not to allowed to have caps on the number of charter schools permitted to operate, and were not to have barriers that prevented linking student-achievement data with individual teacher information (McNeil, 2009).

Supporters of RTTT. Supporters credited RTTT with revolutionizing the federal role in education and transforming state school reform efforts. McGuinn (2012) suggested that more robust state student-data systems and the adoption of common academic standards and assessments may be two of the most important accomplishments of RTTT. McNeil (2011) claimed the eligibility requirements, independent of the specific state grant proposals, had a major impact on school reform efforts. Although RTTT was technically separate from ESEA, the scoring criteria gave insight into the direction that Obama and his team would take into the next ESEA reauthorization (Manna, 2011).

Criticizers of RTTT. Critics referred to RTTT as the *Race to Nowhere, the Race to the Trough, and Dash for Cash* (Martin & Lázaro, 2011). Ravitch, as cited in Martin

and Lázaro (2011) claimed that (a) the inadequate systems of NCLB would still be applied, but on a tougher scale; (b) teacher evaluations would be based on students' scores, and would promote an educational system that focused on scores instead of learning; and (c) the federal government's role in education would become a more important function instead of becoming more decentralized. Levine and Levine (2012) stated that without defining quality, and despite lacking good evidence of their superiority, RTTT wanted more charter schools. They later suggested the message was not about providing support to struggling schools, but about issuing threats of school and job loss. In areas that were heavily populated with minorities, many schools were closed without considering the wishes of the community, or the negative impacts on students (Levine and Levine, 2014).

Teacher and principal evaluation. RTTT was responsible for many local and state policies and laws that resulted in the development and implementation of rigorous teacher and principal evaluation systems. According to Coggshall, Rasmussen, Colton, Milton, and Jacques (2012) evaluation systems were required to use multiple measures to assess educator effectiveness, with at least one of the measures utilizing data on student growth. The intent of the evaluation system was to help fulfill the Obama administration's priority of ensuring great teachers and leaders in our nation's schools, and to assist with making personnel decisions. The RTTT grant winners were required to use teacher and principal evaluation results to determine relevant coaching, induction, and/ or professional development decisions (U.S. Department of Education, 2010b). Local districts were required to conduct annual evaluations of teachers and principals that included timely and constructive feedback, and provided teachers and principals with

data on student growth for their students, classes, and schools (U.S. Department of Education, 2010b, Sec. D [2] iii). According to Geier (2014) addressing the quality of K-12 educators was the first goal of RTTT. While quality pre-service training and appropriate professional development had the ability to improve educational quality, there was a strong belief that valid and reliable administrator and teacher evaluation standards were lacking. While NCLB (2002) required a *highly qualified* educator in every classroom, many believed that highly qualified did not necessarily mean highly effective.

A blueprint for reform. On September 24, 2009, Secretary Duncan delivered a speech entitled *Why We Can't Wait* (Duncan, 2009). He outlined his goals for reform and reaffirmed the need to promote educational excellence, encourage high academic standards, strengthen the field of education, support struggling schools, close the achievement gap, and reduce the dropout rate (Whilden, 2010). Included in his comments were criticisms of NCLB (2002). Duncan (2009) stated, "The biggest problem with NCLB is that it doesn't encourage high learning standards. In fact, it inadvertently encourages states to lower them. The net effect is that we are lying to children and parents and telling kids they are succeeding when, in fact, they are not. We have to tell the truth and we have to raise the bar" (para. 22). Six months later, on March 13, 2010, the Obama administration's ESEA reauthorization, *A Blueprint for Reform,* was released.

According to Manna (2011) the blueprint built on the expectations and assumptions of RTTT, and identified five priorities: (a) college and career-ready students, which included leveraging previous work known as the *Common Core State Standards Initiative* (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010); (b) great teachers and leaders in every

school, which included designing performance evaluation systems that analyze educator effectiveness relative to growth in student achievement (value-added); (c) equity and opportunity for all students, which would ensure that all students were included in state accountability systems, and high and low-poverty schools would have comparable resources; (d) raise expectations and reward excellence, which included continuing the RTTT initiatives for states, and supporting effective school choice; and, (e) promoting innovation and continuous improvement, which included the consolidation of several individual, federal education programs into fewer, larger programs with a more flexible funding stream, and reducing the achievement gap and ensuring student safety through community and parental engagement (U.S. Department of Education, 2010a).

ESEA flexibility waiver. While waiting for Congress to reauthorize ESEA, which was due in 2007, the U.S. Department of Education released the ESEA Flexibility document on September 23, 2011 (U.S. Department of Education, 2012). The waiver was to allow states the opportunity to request flexibility from certain provisions of the No Child Left Behind Act of 2001 (NCLB), authorized by the provisions of the law itself, and to move forward with state and local reform that would improve academic achievement and increase the quality of instruction for all students (U.S. Department of Education, 2012). States choosing to apply were required to describe how they would ensure that local districts would implement the three principles for improving student academic achievement and increasing quality instruction: (a) college and career-ready expectations for all students, which served as the basis for what all students are taught to ensure that every student graduates from high school college and career-ready; (b) state-developed differentiated recognition, accountability, and support, which required states to

implement targeted and differentiated accountability systems, provide rigorous supports and interventions to the lowest-performing schools and schools with the lowest graduation rates, and identify and support low-achieving students based on need; and (c) supporting effective instruction and leadership, which required states to implement teacher and principal value-added evaluation and support systems that provide teachers and principals with the feedback and support they need to improve their practices and increase student achievement (U.S. Department of Education, 2013).

Oregon's Preparation for ESEA Flexibility Waiver Approval

In 2011 the Oregon Legislature passed two Senate bills (SB) to further the goal of improving educator effectiveness. SB 290 (Or. 2011) required the Oregon Department of Education (ODE) to adopt core performance standards for teacher and administrator evaluation. In addition to setting standards for teachers and administrators, ODE also stipulated that: (a) evaluation systems were to be designed collaboratively with teachers and their exclusive bargaining representatives; (b) evaluation systems were to align with the model core teaching standards by July 2013; and (c) multiple sources of data, including value-added assessment, must be used to measure teacher and administrator performance on standards, including a growth process that supports professional learning and collaboration (Oregon Department of Education, 2012). The legislature's second bill, SB 252 (Or. 2011), authorized district collaboration grants that supported integrating performance evaluation systems with career pathways, new compensation models and research-based professional development; support of the labor association, school board and superintendent was required for all grant applications.

The Oregon framework for teacher and administrator evaluation. In response to SB 290 (Or. 2011) Oregon adopted nationally developed standards for teachers and administrators, and required that districts and associations work collaboratively to design new performance evaluation systems. In 2011, Oregon Administrative Rules (OAR) and the Oregon Framework for Teacher and Administrator Evaluation and Support Systems were adopted by ODE in order to obtain approval for its flexibility waiver application (Olney, 2012). ODE (2012) described the Oregon Framework for Teacher and Administrator Evaluation and Support Systems as a fair and equitable system to measure teacher and leader effectiveness. They stated it was a valid and reliable system that would meaningfully differentiate performance and measure teachers' and principals' contributions to student learning and growth toward academic goals and learning standards. The five required elements were: (a) standards of professional practice, (b) differentiated performance levels, (c) multiple measures to evaluate effectiveness based on state board adopted core teaching and administrator performance standards, (d) annual evaluation and professional growth cycles, and (e) aligned professional learning.

Oregon's educator evaluation system. The Oregon Department of Education's educator evaluation system is intended to promote professional growth for teachers and administrators based on the adopted standards of professional practice and meaningful measures of effectiveness (2013). Evaluations require multiple measures of performance and must include evidence of (a) professional practices, (b) professional responsibilities, and (c) student learning and growth (value-added). Professional practices and professional responsibilities are evaluated using the administrative evaluation rubric that aligns to the Interstate School Leaders Licensure Consortium (ISSLC) standards (2008).

Student learning and growth is evidence of administrators' contribution to school-wide student learning and growth, and all administrators must establish at least two student growth goals. One goal must be related to student learning and growth using state assessment as a measure (e.g., building-level data on proficiency and growth in reading and math, including all sub-groups). Evaluations must include student learning and growth as a significant factor and include student academic growth data that is both formative and summative; evidence will be used to inform personnel decisions (Oregon Department of Education, 2013).

Oregon flexibility waiver. Following Oregon's flexibility waiver approval in July 2012, an interim accountability system was put in place for the 2012-13 school year, which served as a bridge between the former system and the new system. Implementation of the new system was required for the 2013-14 school year (Oregon School Boards Association, 2013). A pilot study was conducted to test guidelines developed by the state; data from the pilot was used for refining guidelines and the evaluation process. Federal approval was received to continue the pilot during the 2013-2014 school year. In the spring of 2014, ODE submitted amended guidelines to the U.S. Department of Education that proposed a method for incorporating student learning and growth as a significant measure of teacher and administrator effectiveness that is consistent with the requirements of ESEA flexibility (Oregon Department of Education, 2014).

A Review of Value-added Models

To better understand Oregon's ESEA Waiver, one should understand coupling student test scores to administrator and teacher evaluation. The federal ESEA Flexibility Waiver process required states to focus on educator effectiveness, with the desire to hold

principals and teachers accountable for the academic results of all students (U.S. Department of Education, 2012). Part of that accountability included student assessment data as a measure of effectiveness, which was a key piece of the most current federal requirements of Race to the Top (RTTT). High stakes testing and accountability, and students' and schools' performance on standardized test have become a focal point for both federal and state legislators and policymakers alike (Anderman, Anderman, Yough & Gimbert, 2010). RTTT required participating states to create a new set of assessments aligned to new common academic standards (Whilden, 2010) and to eliminate barriers that prevented linking student-achievement data with individual teacher information (McNeil, 2009). The ESEA waiver continued with the expectations and assumptions of RTTT, with its requirement for designing performance evaluation systems that analyzed educator effectiveness relative to growth in student achievement, and its requirement for all students to be included in state accountability systems (Manna, 2011). At the state level, data systems were developed to (a) match teachers to students and (b) track student test scores from year to year through the use of standardized test scores. Oregon required its educator evaluations to include student learning and growth as a significant factor, and to include student academic growth data that is both formative and summative (Oregon Department of Education, 2013). Importantly, student scores from statewide testing data must be used as one of the measures of educator effectiveness for all Oregon administrators, and for teachers who teach in state-tested subject areas.

Linking student growth to administrator and teacher evaluation. Amrein-Beardsley (2008) referred to the methods of analyzing students' gains, growth in scores, or the amount of knowledge added from year to year as they progress through school as

value-added models (VAM). Lipscomb, Teh, Gill, Chiang and Owens (2010) described VAM as the statistical methods used to estimate educators' contributions to student achievement, and stated that while different VAM calculations may vary in their details, it has a general approach that relies on information about individual students' achievement in other years and/or academic subjects. They described a typical VAM as including an outcome measure, a baseline (or prior year) measurement of the outcome, control variables and teacher or principal variables. Despite these definitions of VAM, Franco & Seidel (2014) stated that as a result of the variety of value-added models being implemented, there was *not* a universal definition for the term value-added. Instead, they suggested the term was used by states to refer to an alternative way to provide feedback on educational effectiveness based on analysis of individual student growth over time.

VAMs for teachers. As noted earlier, value-added calculations are specialized growth models designed to analyze student achievement progress over time, and to understand the amount of student progress that can reasonably be attributed to working with a particular teacher, and in a particular school building, relative to other teachers and schools (Franco & Seidel, 2014). Lipscomb, et al, (2010) described a central principle of VAM as teachers not being held responsible for their students incoming achievement, but instead being evaluated based on how much they contribute to their students' learning. Agreeing, Amrein-Beardsley (2008) stated that VAMs helped to evaluate the knowledge that teachers, schools and school districts add to student learning as they progress through school. Anderman et al. (2010) cautioned that although VAMs are becoming very popular, questions have emerged regarding whether existing state achievement tests can even accurately measure growth across consecutive grades. Armour-Garb (2009)

suggested that while the idea of tracing student performance to individual teachers may seem simple, value-added models are quite technically complicated. She reported that the use of value-added data has created controversy in states where the data is being used in high-stakes decision-making, such as teacher tenure and merit pay. And, in their literature review of 21 studies, Lipscomb, et al. (2010) found while many states were using VAM, they were using the data in a variety of ways; Louisiana was using the data to assess the effectiveness of their teacher education programs, New York was using it to help teachers improve, and in Tennessee it was being used to study the distribution of effective teachers in high poverty schools.

VAMs for administrators. Accountability mandates are tying the definition of effective schools to improved student performance based on student achievement and are holding the instructional leaders of schools accountable for their impact on student achievement (Parylo, Zepeda &Bengston, 2012). Haar (2004) concluded school principals are regarded as key factors in the success of school change and improvement efforts, and are held accountable for teacher performance and school success. Jacobson and Bezzina (2008) suggested that the direct link between principal and student achievement is much more difficult to identify than the indirect influence on school success, teacher performance and instructional quality. The struggle to determine a principal's true contributions to student outcomes from the influence of other school-related factors is a key challenge to calculating principal VAM (Lipscomb et al., 2010). When attempting to address the direct impact that principals have on student achievement, it is important to consider factors beyond their control that may impact student performance (Lipscomb, Chiang & Gill (2012). Those factors may range from teacher to

student to community variables that could positively or negatively influence student test scores that are aggregated to the building level. While teachers have direct instructional contact with students, principals can influence student achievement only indirectly (Lipscomb et al., 2012). According to Lipscomb et al. (2012) estimating the effectiveness of the principal's school is a natural starting point for estimating principal effectiveness that is complicated by circumstances beyond the principal's control, including the pre-existing abilities of the school's teachers. In Franco & Seidel's (2014) review of value-added approaches and the impacts that building-level factors may have on the value-added ratings of teachers, they found that some VAM models involved the use of multivariate statistical methods that took into account socioeconomic status, prior testing results, and student factors such as race, gender ethnicity, native language and mobility, with some that also included the effects of previous teachers. Papay (2010) warned that teacher value-added estimates are sensitive to many characteristics of the tests on which they are based, and policymakers and practitioners who wish to use VAM estimates to make high-stakes decisions must think carefully about the consequences that the differences may have in producing teacher effectiveness ratings. Those specific teacher concerns are then magnified when aggregating all of the teacher's classroom scores to the school building level, which is then utilized to evaluate the principal. Lipscomb et al. (2012) reported that VAMs can be informative tools for identifying highly effective and highly ineffective teachers and schools. However, they also concluded that it is important to consider the limitations of VAMs from factors beyond their control when applying them to large-scale evaluation systems, especially for principals.

Summary and Study Context

This paper reviewed the history of the federal government's primary education law, the Elementary and Secondary Education Act (1965) and the presidential mandates and reauthorizations that have occurred over the past 50 years. While the goals of improving academic outcomes for all students, decreasing the achievement gap, and increasing educator effectiveness and accountability continue to be at the forefront of each administration's education agenda, my review showed that a gap in the research exists on how to best measure administrators' impact and effectiveness on student learning and growth.

Oregon's guidelines, aligned to state legislation and the ESEA Waiver, as illustrated in Figure 2, required districts to incorporate student learning and growth as a significant factor in the overall performance rating of teachers and administrators, and required using student test scores aggregated at the building level through a process called VAM. While the state does not support the use of standardized assessment data as the sole measure of student learning, nor does it use measures of student growth as the sole component on which to base evaluation, both of these components are a requirement of every building administrator's summative performance evaluation. According to the Oregon Framework for Teacher and Administrator Evaluation and Support Systems (Oregon Department of Education, 2014) every building administrator will establish at least two student learning and growth goals. One goal must use Oregon's state assessment data, and must demonstrate evidence of the administrator's impact on the academic growth of all students, regardless of socio-economic status, language, and family background, contributing to overall student success. The second goal may use

measures from statewide assessment data or from district-wide or school-wide measures.

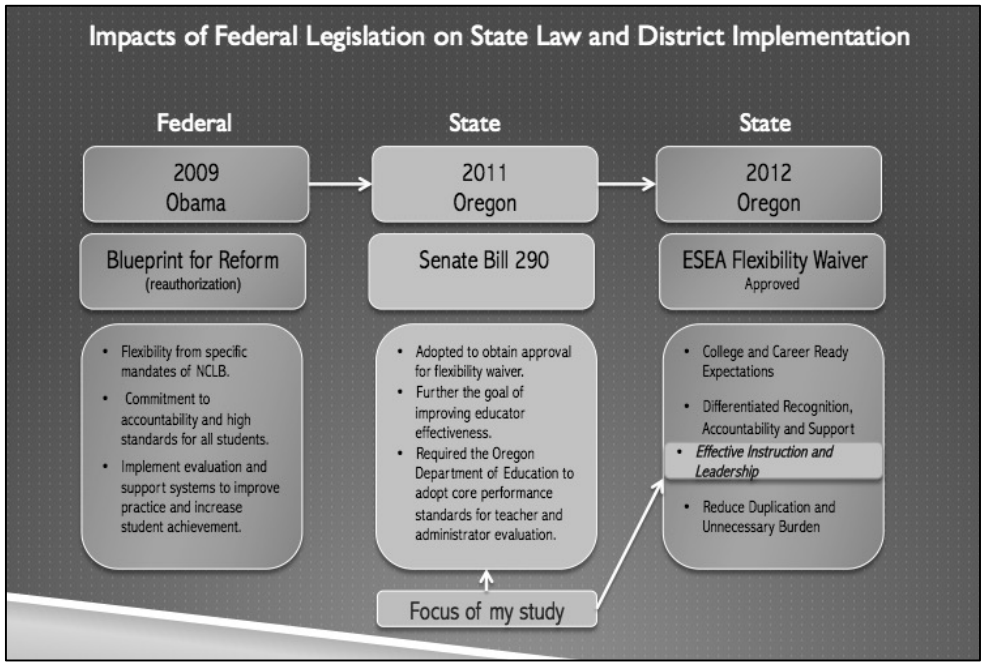

Figure 2. Legislative connection between federal educational reform and Oregon.

For my research, I further investigate which of the two above requirements requiring the use statewide (summative) or school-wide assessment data (formative) better reflects student growth for principals. My findings inform the Federal requirements that are stipulated in state law and must be carried out through a district plan. Again, for my research, I analyze the use of summative versus formative test scores aggregated at the building level to assess which assessment type more effectively measures a building principal's impact on student achievement.

Research Questions

As a result of this gap, my research questions focus on whether large-scale test scores or curriculum-based measures more accurately reflect an elementary principal's ability to impact student learning and growth in reading. My questions are:

1. What is the relationship amongst (a) easyCBM Reading growth metric, (b) OAKS-R Reading growth metric, and (c) elementary principals' summative performance evaluation scores?
2. What is the relative predictive nature of easyCBM Reading growth metrics and the OAKS-R Reading growth metric in relation to the district-wide elementary principals' summative performance evaluation scores?
3. What is the unique contribution of the non-academic predictor variables (demographic risk factors) of (a) attendance, (b) English Language Proficiency (ELP), (c) Free and Reduced Meals (FARMS), (d) percent Other-than-White, and (e) Special Education to the elementary principals' summative performance evaluation scores?
4. Can visual representations depict the academic and non-academic variables?

CHAPTER II

METHODS

Based on previous research, it is important to understand which academic measurement system (formative or summative) best predicts the principal's observation ratings of teachers. My study analyzed extant data from a convenience sample obtained from (a) district-wide elementary principal summative performance evaluation scores, (b) district-wide fourth and fifth grade, fall and spring, reading curriculum-based assessment scores from easyCBM, and (c) 2013 and 2014 spring reading scores from the Oregon Assessment of Knowledge and Skills (OAKS-R).

Research Design

My study utilized a non-experimental, descriptive research design that employed both *t*-test and multiple regression analyses to determine which assessment and/or non-assessment (demographic) factors account for the differences between principals' summative evaluation scores. I used an alpha value of .05 as the cutoff criteria for all tests of statistical significance.

Statistical Analysis

To answer Research Question One, I created a simplified growth metric for analyzing easyCBM Reading and OAKS-R Reading. For easyCBM Reading, I calculated the growth metric using the following formula:

$$CBM\ Growth\ Metric\ = \frac{Spring\ CBM\ Score - Fall\ CBM\ Score}{Grand\ SEM}$$

The CBM growth metric was percentage of the number of students within a principal's school that grew more (or less) than one standard error of measurement (SEM). For OAKS-R Reading, the growth metric was calculated using the following formula:

$$OAKS\ Growth\ Metric\ =\ \frac{2012.13\ OAKS\ Reading\ Score - 2013.14\ OAKS\ Reading\ Score}{Grand\ SEM}$$

The OAKS-R growth metric was a percentage of the number of students within a principal's school that grew more (or less) than one SEM. After calculating each student's growth metric, I used *t*-tests to determine whether significant differences exist between the OAKS-R and easyCBM Reading growth metrics.

For Research Questions Two and Three, I used multiple linear regressions to determine which measures/variables accounted for the most variance in principal ratings. The measures used in this study were the (a) predictor variables consisting of easyCBM Reading growth metric, the OAKS-R Reading growth metric, and student non-academic variables, and (b) a criterion variable consisting of principals' summative evaluation scores. The student non-academic predictor variables (demographic risk factors) were (a) attendance, (b) English Language Proficiency (ELP), (c) Free-and-Reduced Meals (FARMs), (d) percent Other-than-White, and (e) Special Education.

Setting, Participants, and Demographics

I conducted this study using data from 12 elementary schools in a city in the state of Oregon with a population of approximately 58,000 residents, with an overall student population of 11,018 in kindergarten through Grade 12. The participants in this study consisted of all fourth and fifth grade students in the district (n = 1,549) and all elementary school principals (n =12).

The district's population has a total minority population of approximately 35.9%. More specifically, 67.9% of the students in the district are white, 1.4% are Black, 19.8% are Hispanic, 1.4% are Asian/Pacific Islander, 1.6% are American Indian/Alaskan Native, and 5.2% are Multi-Ethnic (Oregon Department of Education, 2012). Based upon the

district's free and reduced data 63.2% of the student population qualify for free or reduced lunches (Oregon Department of Education, 2012).

The district's principals consist of nine women, eight Caucasian and one Pilipino, and three Caucasian men. Three principals have been an administrator between one and five years, one principal has between six and 10 years of experience, and seven principals have between 11 and 15 years of experience.

Measures

My study dependent measures included: (a) all 2013 and 2014 fourth and fifth grade reading scores from OAKS-R; (b) all fourth and fifth grade easyCBM reading scores from the 2013-2014 fall and spring assessments (Multiple-choice Reading Comprehension [MCRC], Passage Reading Fluency [PRF], and Vocabulary [VOC]); and (c) summative principal performance evaluation scores. I collected demographic data specific to the fourth and fifth grade sample student populations.

Principals' summative evaluation scores. The Administrator Growth and Evaluation System's summative performance evaluation score consists of three sub-scores: (a) *professional practice*, which counts for 60% of the summative score; (b) *professional responsibilities*, which counts for 20% of the summative score; and (c) *student learning and growth*, which counts for 20% of the summative score (see Appendix A). However, for the purposes of my research the principal summative evaluation score combined the first two categories (*professional practice* and *professional responsibilities*), because the third category (*student learning and growth*) is the focus of my research. *Professional practice* and *professional responsibility* measures are scored using a four-point, continuous scale rubric based upon the six Educational

Policy Leadership Standards for School Leaders, developed by the Interstate School Leaders Licensure Consortium (ISLLC). Evidence of *student learning and growth* is collected through multiple sources, and is also scored on a district-designed, four-point, continuous scale rubric. In collaboration with their supervisor/evaluator, the Oregon Department of Education requires that administrators establish at least two Student Learning and Growth goals each year (see Appendix B); one goal must be related to student learning and growth using the OAKS-R as a measure (e.g., building-level data on proficiency and growth in reading and math, including all subgroups).

OAKS-R. For my summative evaluation, I used data from (a) the OAKS-R (Reading), a large-scale, summative assessment, which is given every year to all fourth and fifth grade students; (b) easyCBM Reading, a formative assessment that is given seasonally to all fourth and fifth grade students in the fall and spring; and, (c) Synergy, a data collection system used to collect school factor information. Summative assessments are typically used for school accountability, program evaluation, and to estimate groups of students' achievement levels. The OAKS-R test aligns to Oregon content standards and measures the breadth of the standards, which are comprised of a broad range of knowledge and skills, and is not sensitive to the gains in student learning that occur between a few weeks of instruction (Oregon Department of Education, 2012).

The OAKS-R Online Reading/Literature assessment is a criterion-referenced and computer adaptive state test. The Reading/Literature assessment is administered in Grades 3-8 and in Grade 10. Each item was a question or statement that required completion and students selected from four answer choices. The first item is of average difficulty for the specified grade. The assessment algorithm then selects subsequent items

based on the number of items already presented for each content strand, the degree the item maximizes precision in identifying the student's proficiency, and the student's ability based on performance on earlier items (Oregon Department of Education, 2012b).

***OAKS-R reliability*.** Reliability of the OAKS-R Online Reading/Literature assessment was examined through several different approaches (Oregon Department of Education, 2007). Standard error curves indicated reliable scores across the ability range and consistent amounts of error regardless of ethnicity, language proficiency, or special education designations. High achievement classification reliability was reported, ranging from 84% to 99%, with most falling above 90% (Oregon Department of Education, 2007). Another ODE test of reliability examined the likelihood of students earning false-positive or false-negative scores. Testing showed that the number of these false-positive or false-negative scores were extremely low (Oregon Department of Education, 2007). A final test of reliability was around the accuracy of cut scores for individual strands within a particular test. Overall, there is strong reliability with the OAKS-R test.

***OAKS-R validity*.** Multiple types of validity have evaluated the OAKS-R test. First, the OAKS-R test showed construct validity through the use of rigorous standards developed by multiple stakeholders including educators and other citizens, a test-development process whereby questions were vetted by a panel of educators from multiple areas around the state, and ongoing studies to help evaluate the alignment between assessment and instruction (Oregon Department of Education, 2007). Concurrent validity was also addressed. Multiple studies found high correlations ranging from .71 to .82 between students' reading performance on the OAKS-R and their performance on the California Achievement Test, Iowa Test of Basic Skills, the

Northwest Evaluation Association Subject Test, and the Lexile Scale (Oregon Department of Education, 2007).

easyCBM reading. easyCBM is a benchmark, progressing monitoring, and formative reporting assessment tool for Grades K-8. It was designed for use in measuring student achievement in math and reading, and its assessments are aligned to Common Core State Standards. To examine how fluency measures are associated with overall reading comprehension on the OAKS-R, I utilized individually administered, one-minute PRF and VOC, as well as MCRC.

Using fluency CBMs to measure reading skills has become widely used in the past two decades. While there are a myriad of possibilities for measuring growth over time, many of the assessments have not been developed using sophisticated statistical analyses to determine passage difficulty and equivalency that easyCBM employs. The passages to be used in this study were created for the easyCBM website and developed using more advanced statistical techniques than simply comparing means and standard deviations when administered to a group of students. The passages were initially written and revised in an effort to produce 20 alternate forms of grade level passages to be used as progress monitoring and benchmark passages (Alonzo, Park, & Tindal, 2008). The passage developers paid close attention to several criteria while writing the passages. Specifically, each passage (a) tells a story, (b) does not contain dialog, and (c) stands alone with no references to other passages. Graduate students in the University of Oregon's College of Education wrote the passages. The passages were reviewed for grammar, sentence structure, and grade-level appropriateness by a university professor who is a National Board for Professional Teaching Standards certified English teacher

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

CHAPTER III

RESULTS

I present results in order by research question, with a short summary following.

Research Question One: Academic Measure Relations

For Research Question One I analyzed the relationship between Summative Principal Ratings and easyCBM and OAKS-R spring academic measures. On a one-to-four rating scale, the mean Principal Summative Rating was 2.96 (*SD* = 0.26). Overall, the Principal Summative Rating scores were poorly correlated to all academic measures. The highest correlation was between the Principal Summative Rating Scores and easyCBM PRF, *r* = .09, or 0.77 percent of the easyCBM PRF score was accounted for by principal ratings. The second highest correlation was between the Principal Summative Rating Scores and OAKS-R Reading, *r* = .06, or only 0.35 percent of the OAKS-R score was accounted for by principal ratings. The third highest correlation was between the Principal Summative Rating Scores and the easyCBM Passage MCRC score, *r* = .05, or 0.25 percent of the easyCBM MCRC score was accounted for by principal ratings. The lowest correlation was between the Principal Summative Rating Scores and easyCBM VOC, *r* = .05, or 0.20 percent of the easyCBM VOC score was accounted for by principal ratings. Descriptive statistics and bivariate correlations for the academic measures are displayed in Tables 1 and 2, respectively.

Table 1

Descriptive Statistics for Academic Measures

Measure	*M*	*SD*	*n*
Principal Summative Rating	2.96	0.26	12
S 13.14 MCRC	13.65	3.57	2140
S 13.14 PRF	141.03	50.68	2366
S 13.14 VOC	17.14	2.86	2249
OAKS-R 13.14 RIT	217.27	19.14	2427

Table 2

Bivariate Correlations for Academic Measures

Measure	Principal Summative Rating	S 13.14 MCRC	S 13.14 PRF	S 13.14 VOC
S 13.14 MCRC	.05*			
S 13.14 PRF	.09*	.52*		
S 13.14 VOC	.05*	.56*	.51*	
OAKS-R 13.14 RIT	.06*	.56*	.60*	.61*

* = correlation is significant at the 0.05 level.

Research Question Two: Reading Growth Indicators

For Research Question Two I analyzed the predictive relation between the easyCBM MCRC, PRF, and VOC scores and the Summative Principal Ratings.

easyCBM MCRC and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean gain for the easyCBM MCRC was 1.90 (*SD* = 3.56). The minimum gain was –10.0, and the maximum gain was 14.0. Figure

3 shows the low correlation between easyCBM Comprehension and the Summative Principal Ratings. The quality of the fit was near zero ($R^2 = 0.02$), meaning about two percent of the easyCBM Reading Comprehension was accounted for by principal ratings. The coefficient of –0.01 demonstrated that for every point gained on the Comprehension Gain Score the Principal Summative Rating dropped 0.01 points (Figure 3).

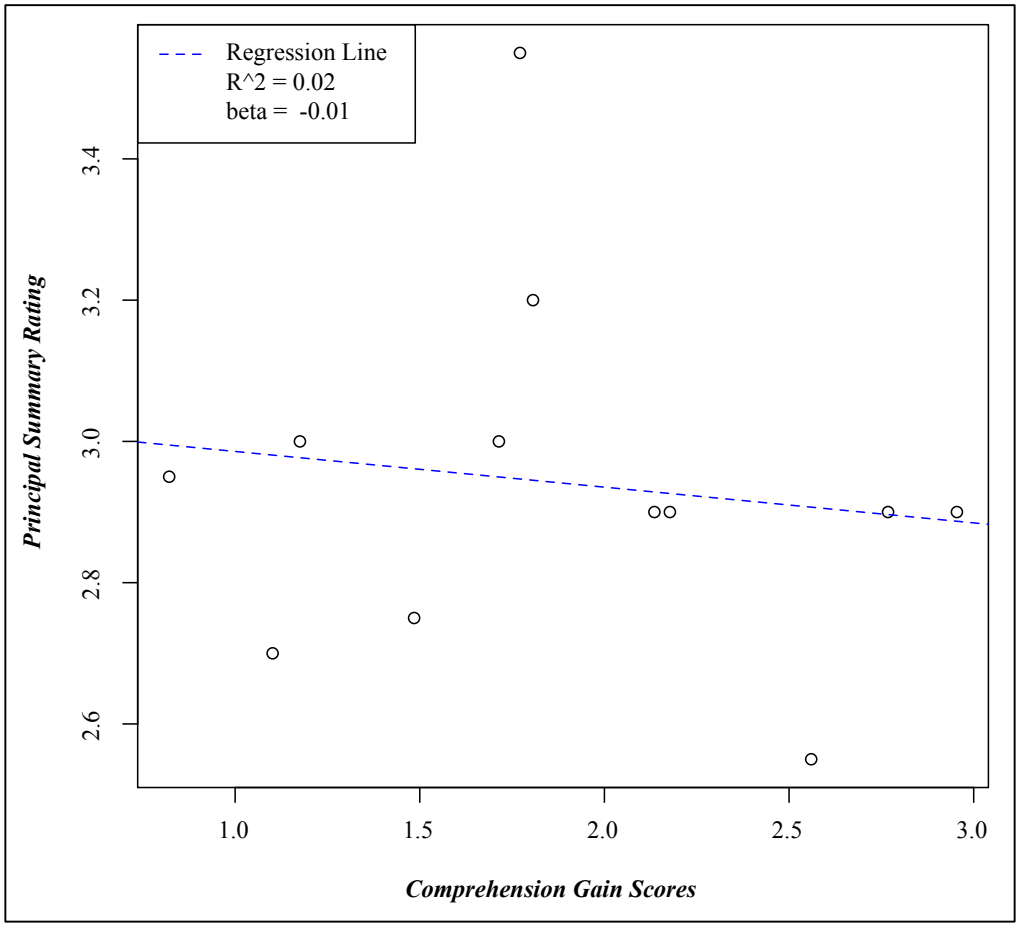

Figure 3. easyCBM MCRC Compared to Summative Principal Ratings.

easyCBM PRF and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean gain for the easyCBM Reading Fluency was 32.15 (*SD* = 20.92). The minimum fluency gain was a –47.0 words per minute; the maximum gain was 114.0 words per minute. Figure 4 shows the low

correlation between easyCBM Comprehension and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.36$, was low moderate, and meant that about 36 percent of the easyCBM Reading Fluency was accounted for by principal ratings. The coefficient of 0.00 demonstrated no slope for the line of best fit. For every point gained on the Reading Fluency Gain Scores, the Principal Summative Rating gained 0.00 points (Figure 4).

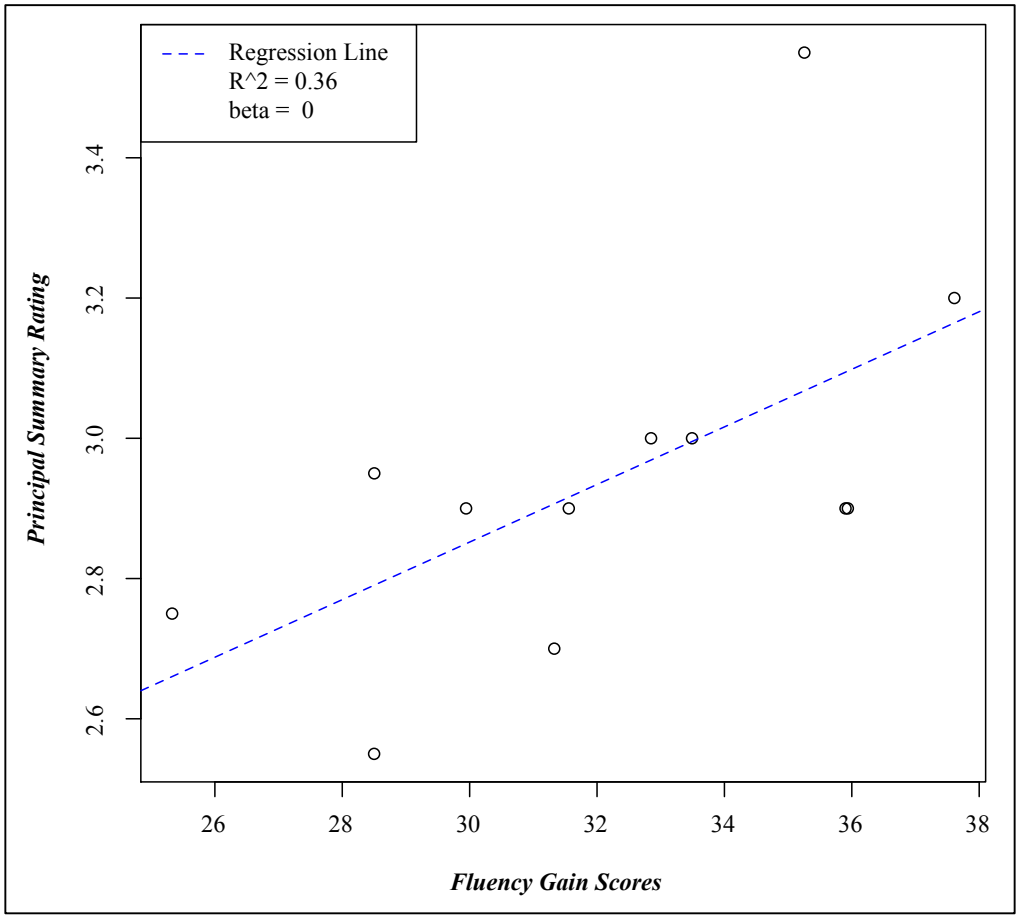

Figure 4. easyCBM PRF Compared to Summative Principal Ratings.

easyCBM VOC and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean gain for the easyCBM Vocabulary was 1.85 (*SD* = 2.90). The minimum vocabulary gain was –8.0; the maximum gain was 17.0. Figure 5 shows the very low correlation between easyCBM Vocabulary and the

Summative Principal Ratings. The quality of the fit, $R^2 = 0.07$ means that about seven percent of the easyCBM Vocabulary was accounted for by principal ratings. The coefficient of –0.02 demonstrated that for every point gained on the Vocabulary Gain Scores, the Principal Summative Rating dropped 0.02 points (Figure 5).

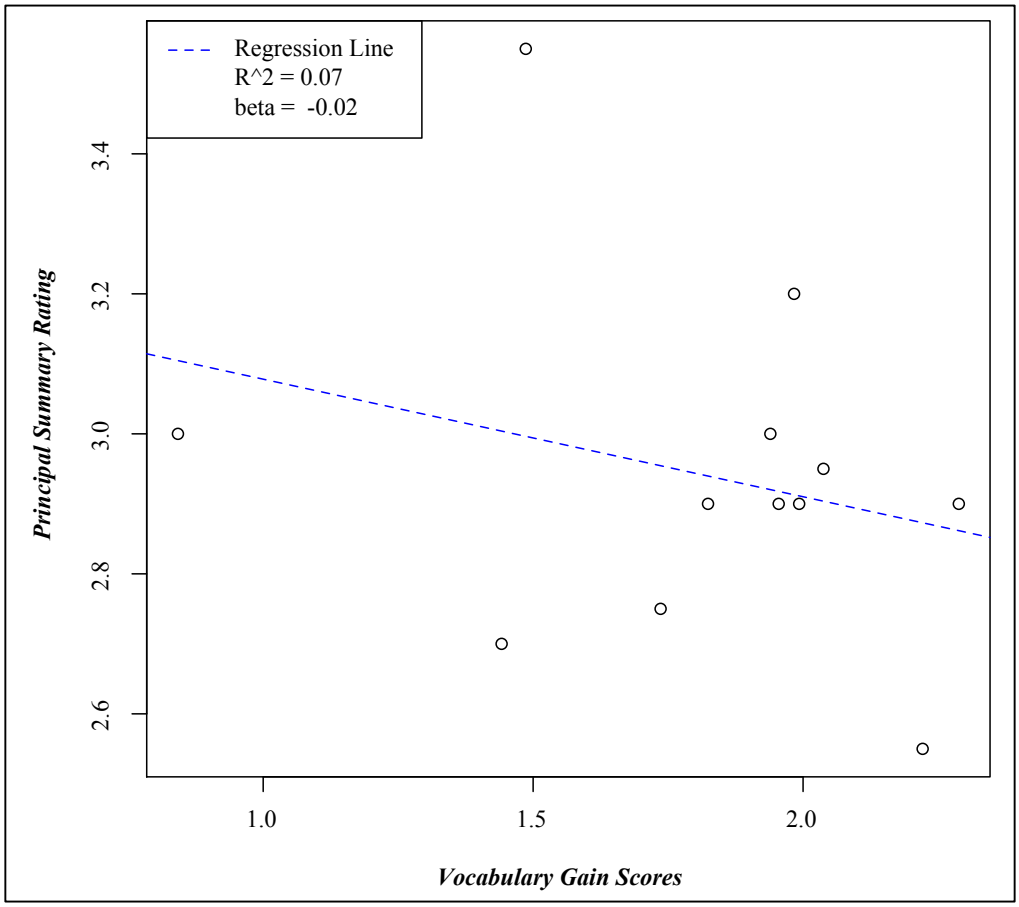

Figure 5. easyCBM VOC Compared to Summative Principal Ratings.

OAKS-R growth and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean gain for the OAKS-R Reading Comprehension was 2.72 (*SD* = 21.87). The minimum RIT point gain was –125.0; the maximum RIT point gain was 145.0. Figure 6 shows the positive correlation between OAKS-R Reading and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.06$

means principal ratings accounted for about six percent of the OAKS-R. The coefficient of 0.01 demonstrated that for every point gained on the OAKS-R Reading Gain Scores, the Principal Summative Rating only gained 0.01 points (Figure 6).

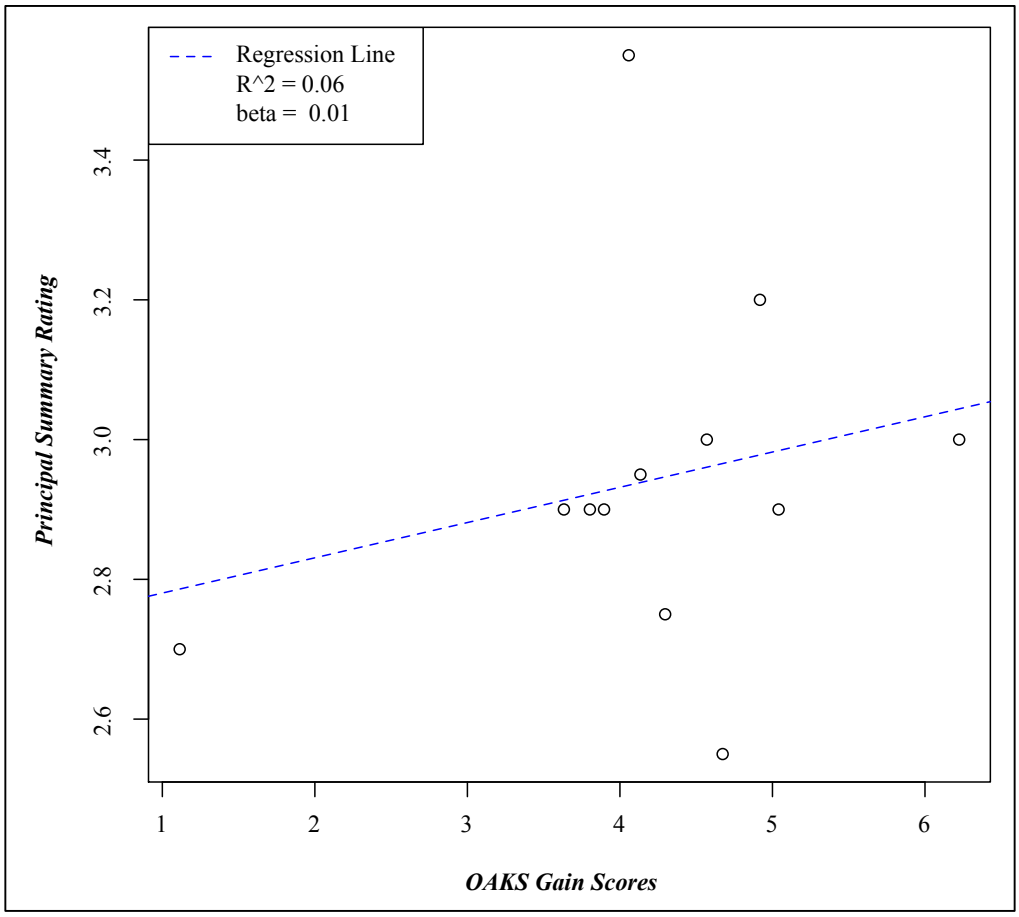

Figure 6. OAKS-R Reading Compared to Summative Principal Ratings.

Research Question Three: Demographic Risk Indicators

Attendance percentages and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean percentage for attendance was 96.08 (*SD* = 5.23). Figure 7 shows the positive correlation between Attendance and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.14$, was low moderate and means that about 14 percent of the Attendance variable was accounted for by principal

ratings. The coefficient of 0.66 demonstrated a positive slope for the line of best fit. For every point gained on the Attendance Score, the Principal Summative Rating gained 0.66 points (Figure 7).

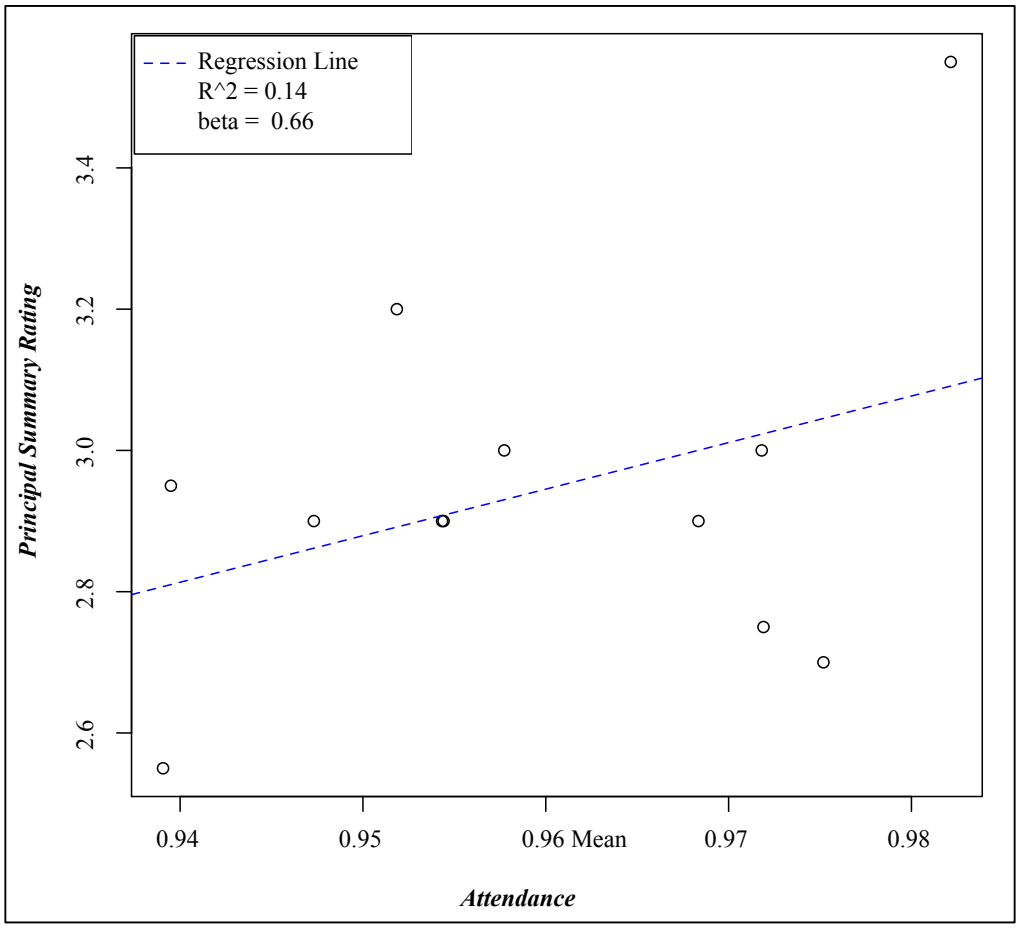

Figure 7. Attendance Percentage by Principal Summary Ratings.

ELP and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean percentage for the ELP was 9.76 (*SD* = 29.68). Figure 8 shows the negative correlation between the ELP variable and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.19$, was low and means that about 19 percent of the ELP was accounted for by principal ratings. The coefficient of –0.18 demonstrated a

slightly negative slope for the line of best fit. For every point gained on the ELP, the Principal Summative Rating lost 0.18 points (Figure 8).

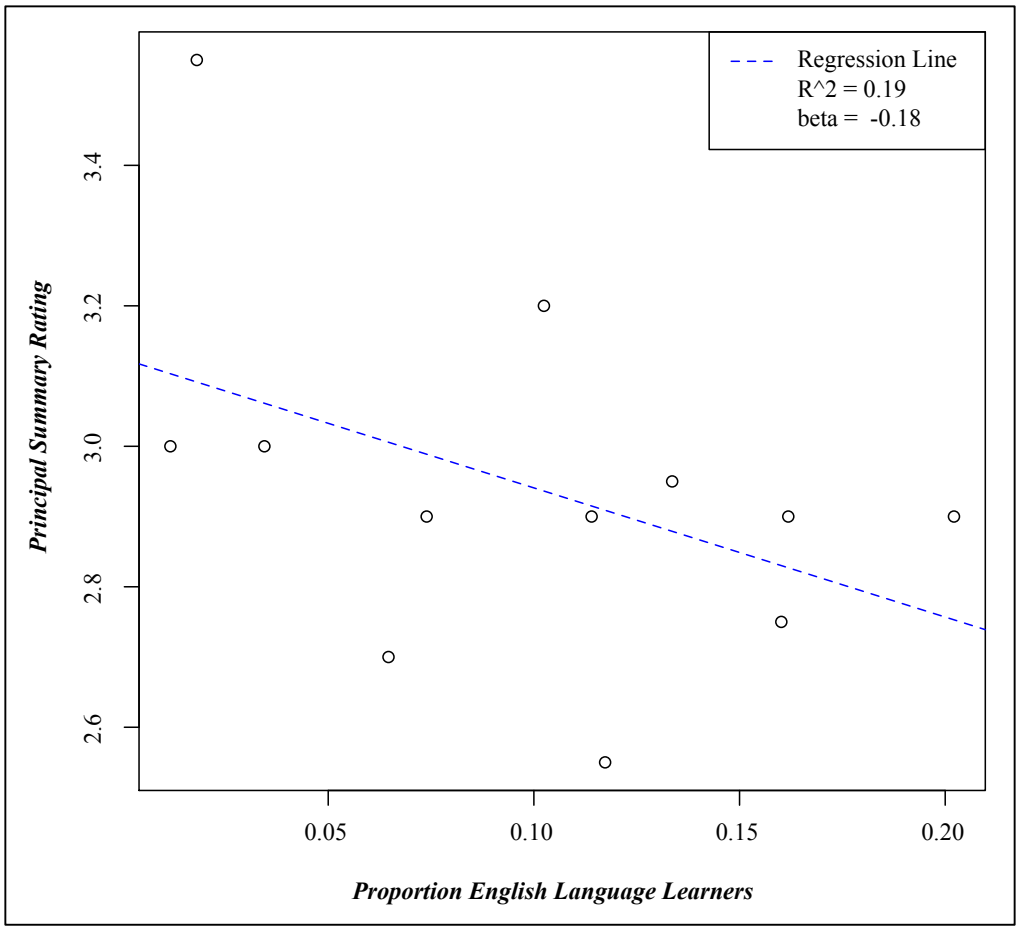

Figure 8. English Language Learner percentage by Principal Summary Ratings.

FARMs and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean percentage for FARMs was 63.26 (*SD* = 48.22). Figure 9 shows the negative correlation between FARMs and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.11$, was low and means that about 11 percent of FARMs was accounted for by principal ratings. The coefficient of –0.05 demonstrated a minor negative slope for the line of best fit. For every point gained on the FARMs, the Principal Summative Rating lost 0.05 points (Figure 9).

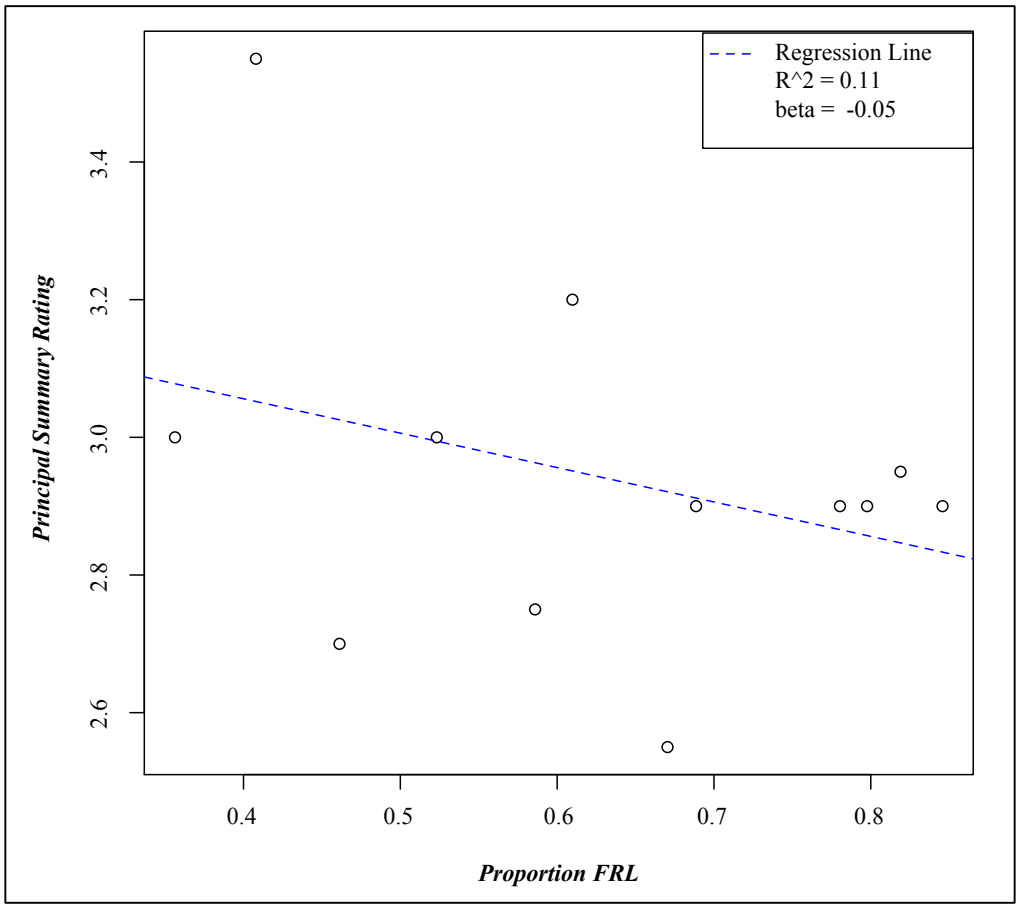

Figure 9. FARMs percentage by Principal Summary Ratings.

Percent Other-than-White and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean percentage for the percentage of students classified as Other-than-White was 32.82 (*SD* = 41.96). Figure 10 shows the negative correlation between Other-than-White and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.11$, was low and means that about 11 percent of the Other-than-White Score was accounted for by principal ratings. The coefficient of –0.11 demonstrated a negative slope for the line of best fit. For every point gained on the Other-than-White, the Principal Summative Rating dropped 0.11 points (Figure 10).

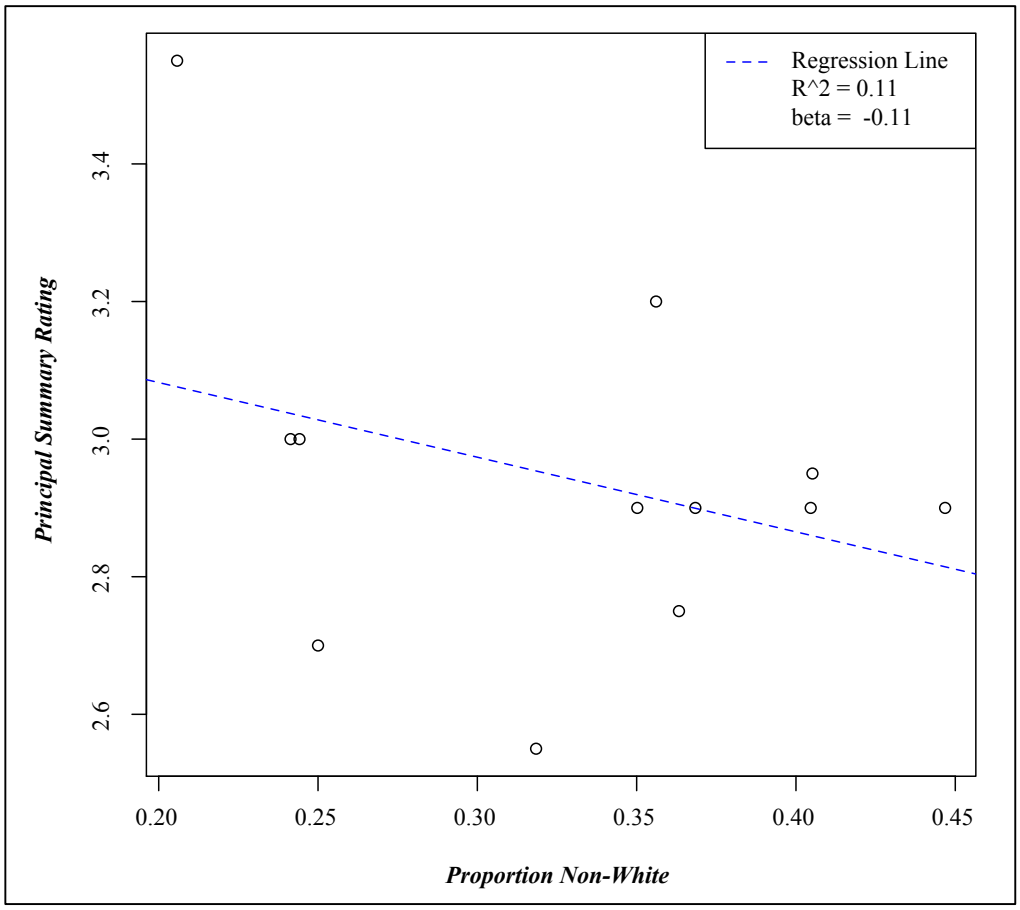

Figure 10. Other-than-White percentage by Principal Summary Ratings.

Special Education and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). The mean percentage for the students in special education was 16.45 (*SD* = 37.08). Figure 11 shows the negative correlation between easyCBM Comprehension and the Summative Principal Ratings. The quality of the fit, R^2 = 0.00, was poor and means that approximately zero percent of the Special Education Score was accounted for by principal ratings. The coefficient of –0.02 demonstrated a slight negative slope for the line of best fit. For every point gained on the Special Education score, the Principal Summative Rating lost 0.02 points (Figure 11).

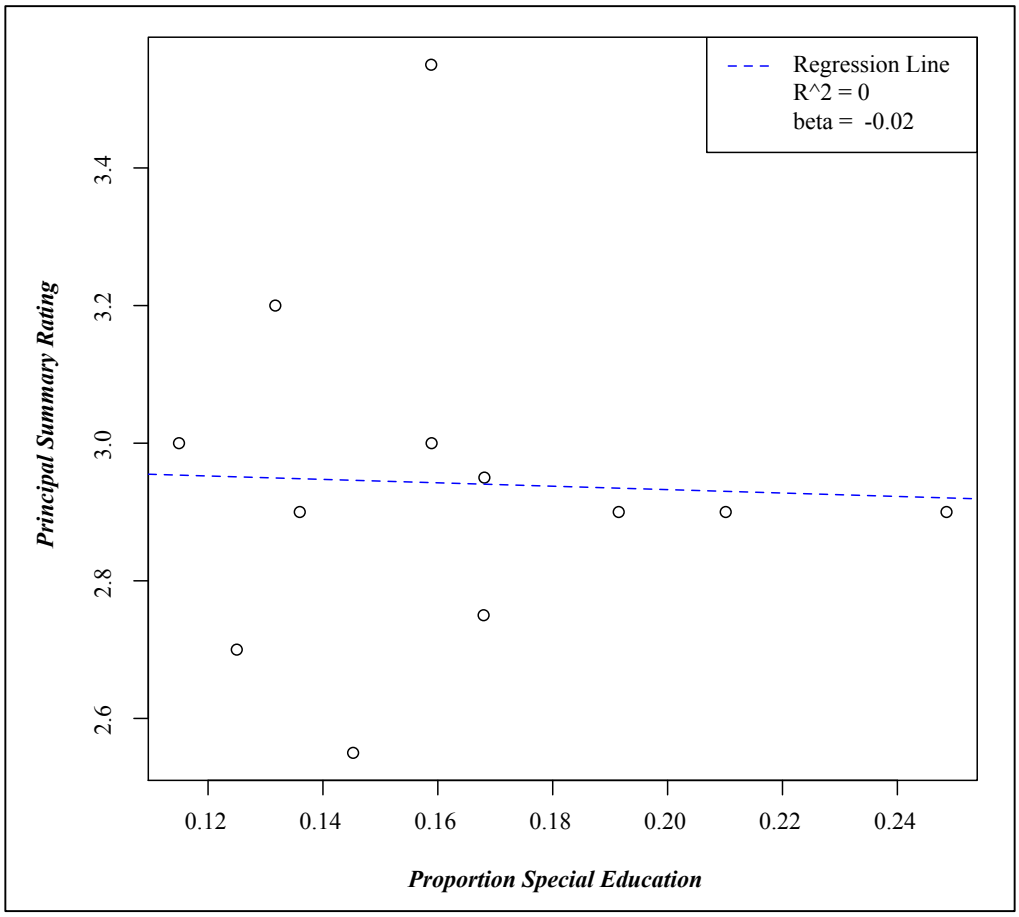

Figure 11. Special Education percentage by Principal Summary Ratings.

PCA risk factors and Summative Principal Ratings. The mean Principal Summative Rating was 2.96 (*SD* = 0.26). Figure 12 shows the negative correlation between the PCA (demographic) Score and the Summative Principal Ratings. The quality of the fit, $R^2 = 0.15$, was low and means that about 15 percent of the PCA Score was accounted for by principal ratings. The coefficient of –0.03 demonstrated a slight negative slope for the line of best fit. For every point gained on the PCA Score, the Principal Summative Rating lost 0.03 points (Figure 12).

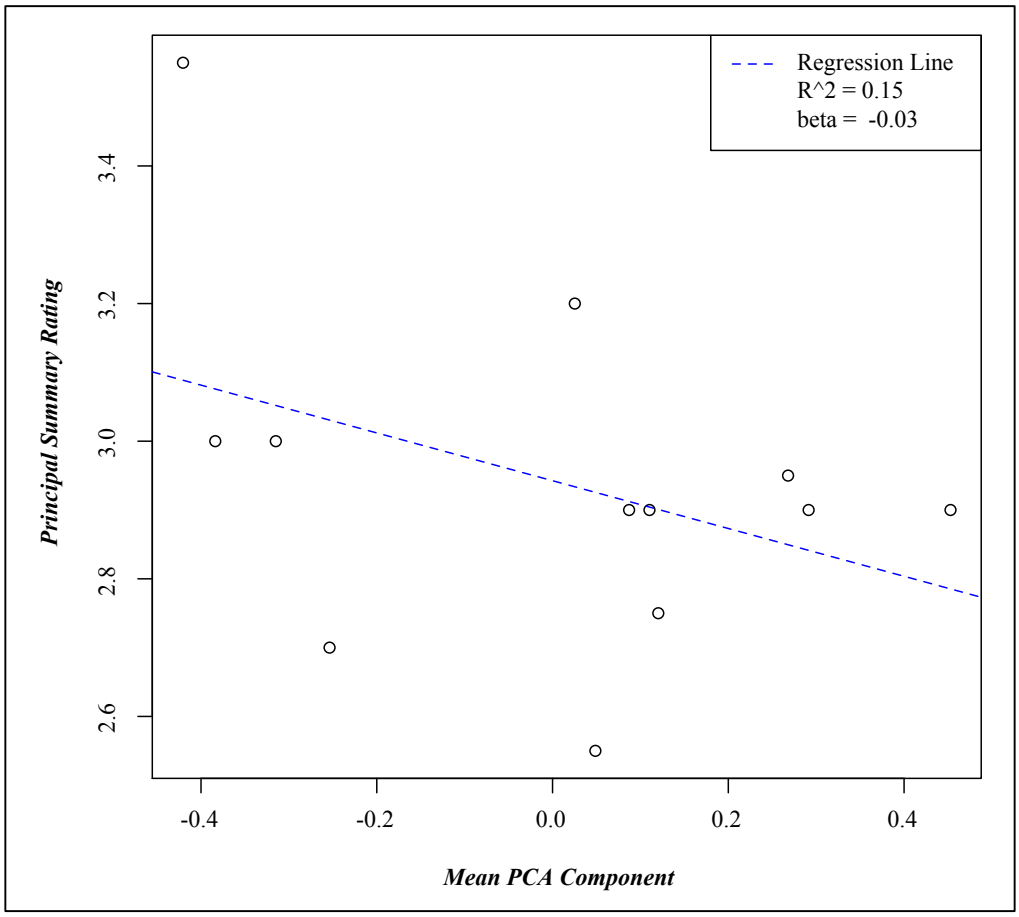

Figure 12. Principle component analysis for risk by Principal Summary Ratings.

Research Question Four: Visual Representation of Variables

Because of the low correlations for Research Question One and the poor quality of fit for Questions Two and Three, I created representations for the academic and non-academic variables in relation to each principal's summative rating and describe these visual patterns, below.

Schools ranked by Summative Principal Ratings and demographics. Figure 13 shows the percentage of (a) FARMs, (b) Special Education, (c) ELP, (d) White, and (e) Attendance. Schools are ranked with School 01 having the principal with the highest summative rating and School 12 having the principal with the lowest summative rating.

Based on the data in the Figure 13, Schools 01, 02, 03, and 04 have the fewest percentage of students with disabilities, the fewest percentage of students who are ELP, while having the highest attendance rates and percentage of White students. Schools 06, 07, and 08, whose principals are ranked sixth, seventh and eighth, respectively, in summative ratings scores, have the highest numbers of students with disabilities, the highest number of students who qualify for FARMs, and the lowest attendance rates. In addition, schools 07 and 08 also have the highest number of non-White and ELP students. Schools 09, 10, 11, and 12, who have the principals with the lowest rated summative evaluation scores, have fewer students with disabilities, fewer ELP students, and fewer students who qualify for FARMs.

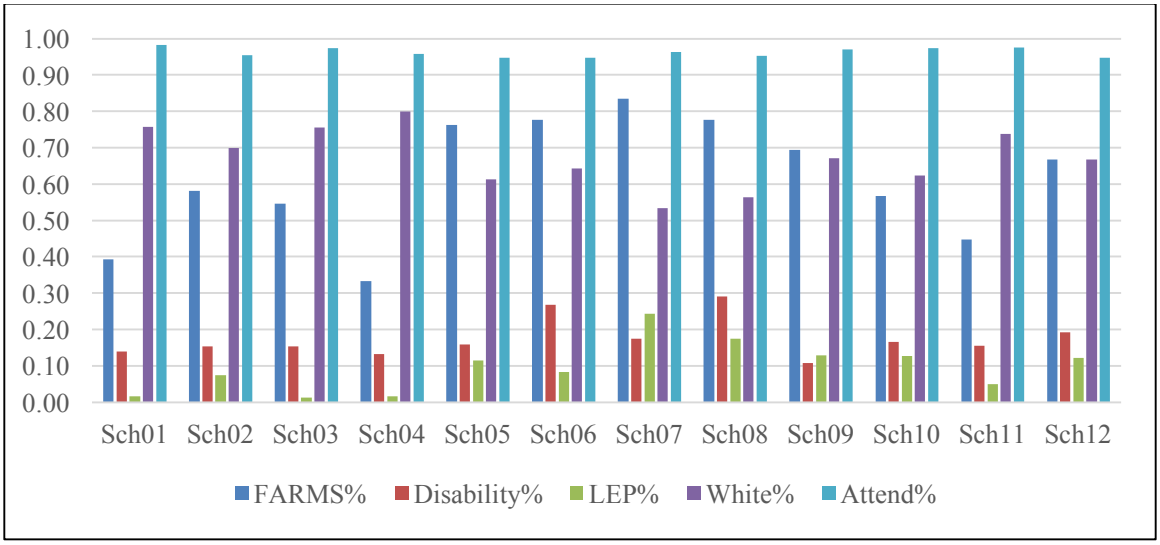

Figure 13. Schools ranked by Summative Principal Ratings and demographic variables.

Another way to view the non-academic variables presented in Figure 13 is through a 100% stacked bar chart. Figure 14 is a 100% stacked bar chart which shows the combined contribution of values of all the non-academic demographic variables as percentages where the combined total for each school is 100 percent. This chart type

displays multiple variables as stacked columns, and the cumulative proportion of each stacked element always totals 100%. The 100% stacked column chart emphasizes the contributions (by school) of (a) FARMs, (b) Special Education, (c) ELP, (d) White, and (e) Attendance to the whole school.

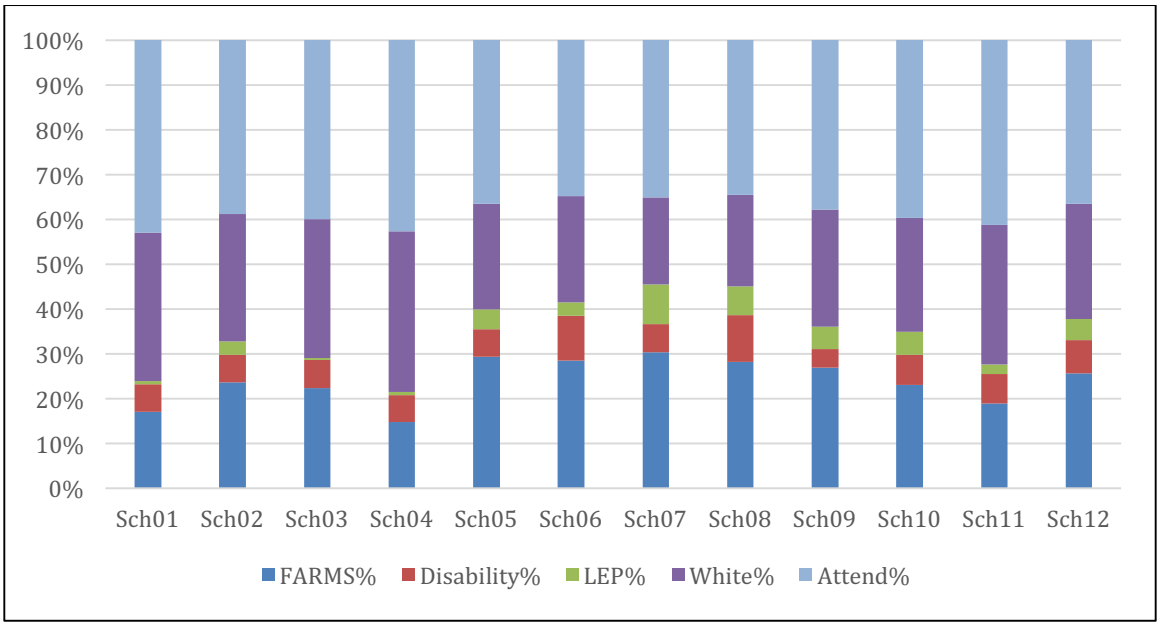

Figure 14. Schools ranked by Summative Principal Ratings and demographic variables (stacked).

Schools ranked by Summative Principal Ratings and PRF. Figure 15 shows the growth of students in PRF on the easyCBM formative assessment from fall to spring of 2013-2014. The schools are rank-ordered with School 01 having the principal with the highest summative rating, and School 12 having the principal with the lowest summative rating. Data indicates School 01, 02, 08, and 09 students scored the highest, with the students in Schools 05 and 06 scoring the lowest.

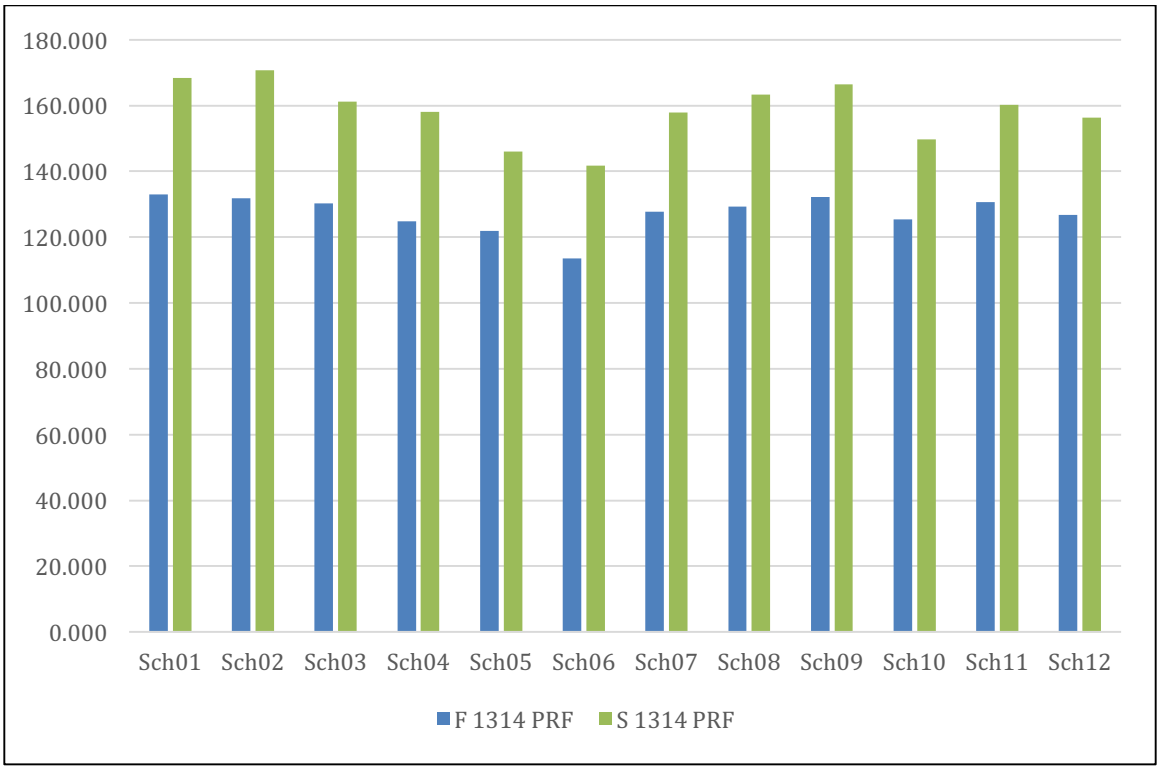

Figure 15. Schools ranked by Summative Principal Ratings and PRF.

While it is notable that the students attending the schools with principals who have the lowest summative rating scores did not have the lowest PRF scores, Figure 16 adds non-academic variables by school to the PRF analysis. When viewing the fall and spring PRF scores through the non-academic variable lens, one can see that the data indicate that students in Schools 01, 02, 08, and 09 scored the highest on their spring PRF scores. Looking at Figure 16 shows that the higher PRF score schools (01 and 02) have the lowest percentage of students with disabilities, a lower percentage of students who are ELP, while having the highest attendance rates and percentage of White students, but School 08 has one of the highest number of non-White and ELP students. School 09 is closer to 01 and 02 with fewer students with disabilities, fewer ELP students, and fewer students who qualify for FARMs.

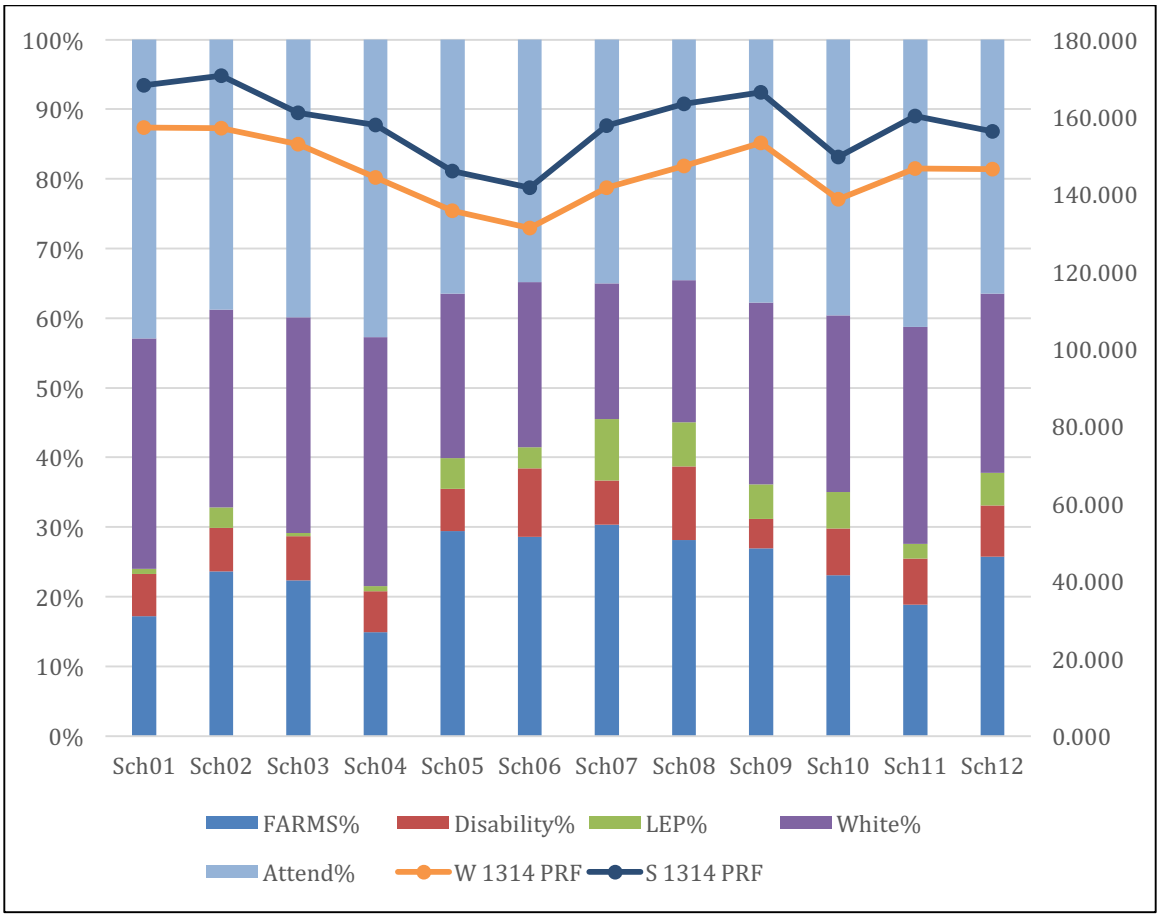

Figure 16. Schools ranked by Summative Principal Ratings and non-academic factors and PRF (stacked).

Schools ranked by Summative Principal Ratings and MCRC. Figure 17 displays the growth of students in MCRC scores on the easyCBM formative assessment from fall to spring of 2013-2014. Again, schools are rank-ordered with School 01 having the principal with the highest summative rating, and School 12 having the principal with the lowest summative rating. These data indicate that the students in School 04 have the highest formative MCRC scores, with the students in Schools 07 and 12 showing the most growth between the fall and spring tests, and the students attending Schools 05 and 11 showing little to no growth.

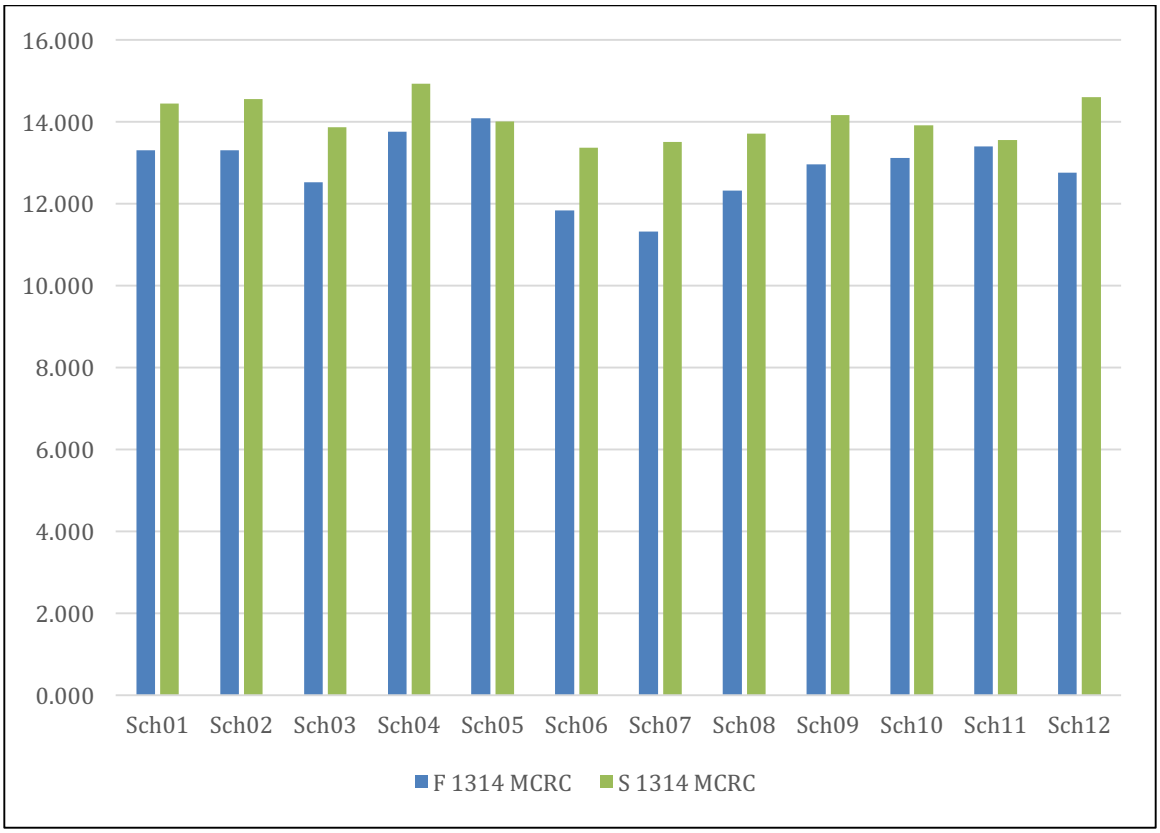

Figure 17. Schools ranked by Summative Principal Ratings and MCRC.

Like the previous two figures for passage reading fluency, students attending the schools with principals with the lowest summative rating scores did not have the lowest multiple-choice reading scores, Figure 18 shows that the students in School 07, which is the school with the highest percent of FARMS and LEP students, as well as the school with the smallest percentage of white students, demonstrated the most growth from fall to spring, with the students in School 12 demonstrating the second highest rate of growth from fall to spring. While having different demographic variables, the students attending Schools 05 and 11, showed little to no growth. It is notable that students in Schools 01 and 02 demonstrated some of the smallest the growth from fall to spring and those

schools had the lowest percentage of students with disabilities and classified as LEP, while having the highest attendance rates and percentage of White students.

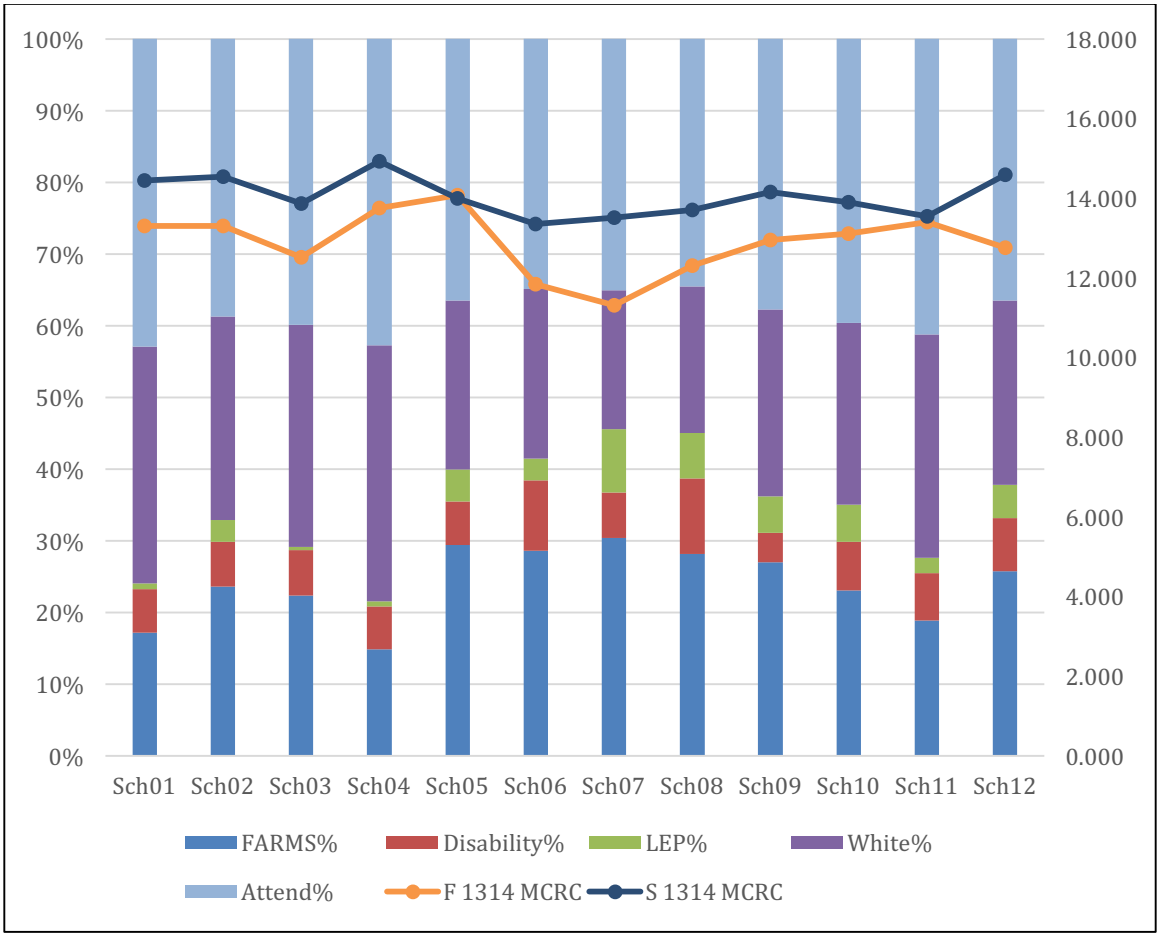

Figure 18. Schools ranked by Summative Principal Ratings and non-academic factors and MCRC (stacked).

Schools ranked by Summative Principal Ratings and VOC. Figure 19 shows the growth of students in VOC scores on the easyCBM formative assessment from fall to spring of 2013-2014. The schools are rank-ordered with School 01 having the principal with the highest summative rating and School 12 having the lowest principal summative rating. These data indicate that the students of Schools 02, 09, 12, and 01, in that order, scored the highest, while the students attending Schools 06 and 07 scored the lowest on

the VOC measure. The students demonstrating the most growth from fall to spring were in School 12. These data reveal that the principal with the lowest summative rating had the students who demonstrated the most growth from fall to spring. Likewise, the principal with the highest summative rating had the students who demonstrated some of the least amount of growth.

Figure 19. Schools ranked by Summative Principal Ratings and VOC.

Figure 20 visually displays the growth of students in VOC scores on the easyCBM formative assessment from fall to spring of 2013-2014 in relation to each school's non-academic factors. Again, data indicates that students of Schools 02, 09, 12, and 01 scored the highest, while students attending Schools 06 and 07 scored the lowest on the VOC measure. The students demonstrating the most growth from fall to spring

were in schools 9 and 12 and 2. As noted earlier, Figure 20 shows that the higher Vocabulary score schools (01 and 02) have the lowest percentage of students with disabilities, a lower percentage of students who are ELP, while having the highest attendance rates and percentage of White students. Moreover, Schools 09 and 12 also have fewer students with disabilities, fewer ELP students, and fewer students who qualify for FARMs. The two low scoring schools (06 and 07) had the highest number of students with disabilities, the highest number of students who qualified for FARMs, and the lowest attendance rates.

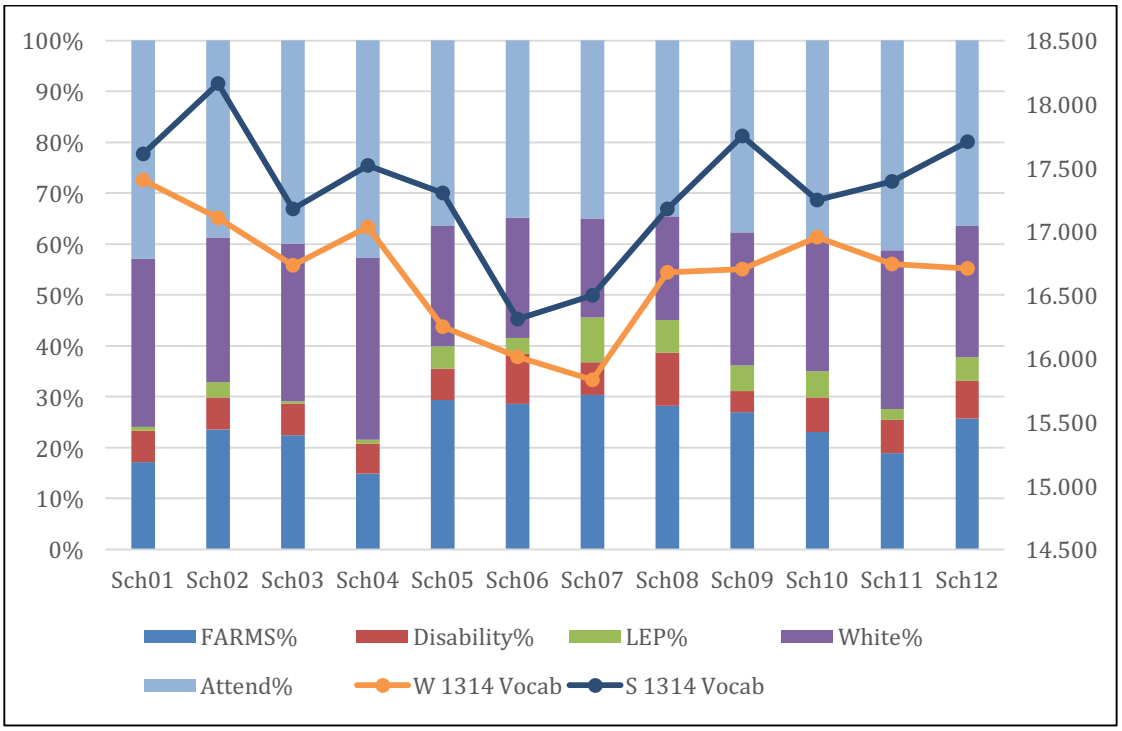

Figure 20. Schools ranked by Summative Principal Ratings and non-academic factors and VOC (stacked).

Schools ranked by Summative Principal Ratings and OAKS-R. Figure 21 shows the growth of students in the large-scale summative assessment OAKS-R scores from the spring of 2013 to the spring of 2014. The schools are rank-ordered with School

01 having the principal with the highest summative rating and School 12 having the principal with the lowest summative rating. These data indicate that the students in Schools 01, 09, and 12 have the highest OAKS-R scores, with School 06 demonstrating significantly more growth than any other school.

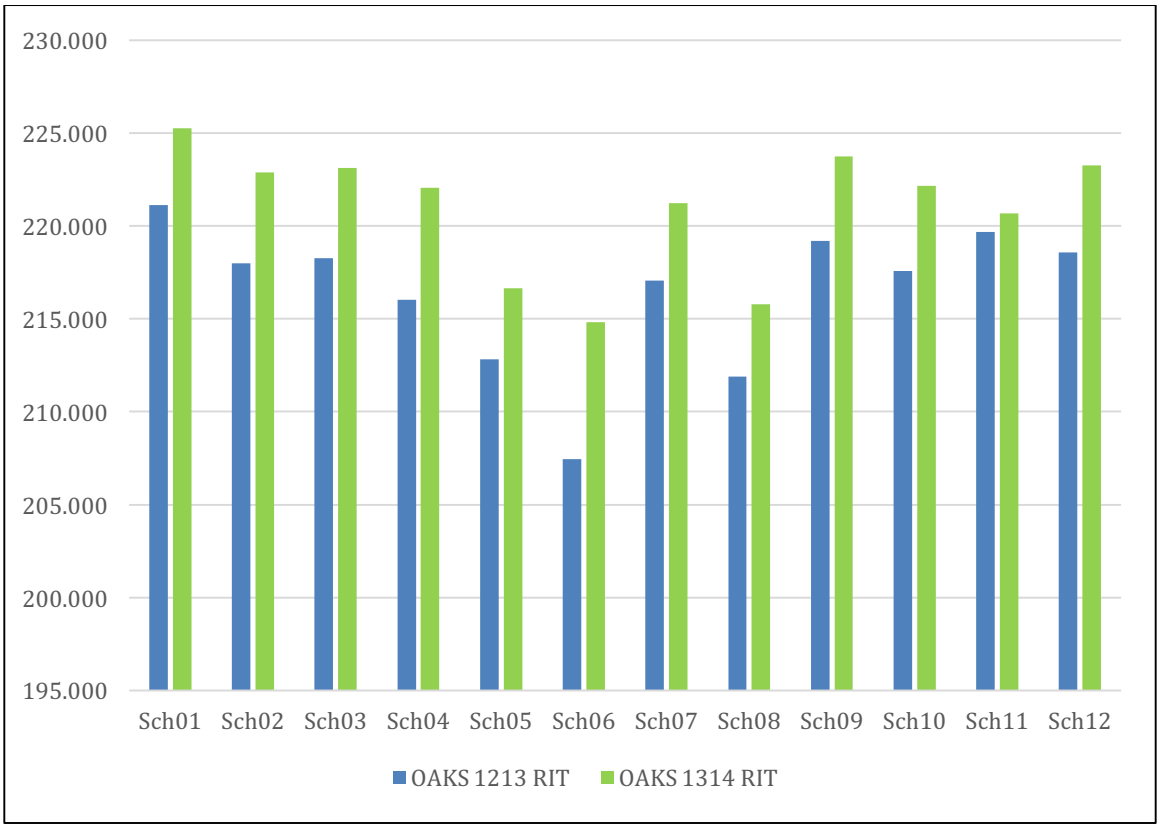

Figure 21. Schools ranked by Summative Principal Ratings and OAKS-R.

Figure 22 shows the growth of students in the large-scale summative assessment OAKS-R scores from the spring of 2013 to the spring of 2014 compared to each school's non-academic factors. Again, data indicates that the students in Schools 01, 09, and 12 have the highest OAKS-R scores from 12-13 to 13-14. However, non-academic (demographic) factors in Figure 22 show that the students in schools 01, 09, and 12 have the highest OAKS-R scores, while school 06 which has the lowest overall scores,

demonstrates considerably more growth than any other school from 12-13 to 13-14. Interestingly, School 01 has one of the the lowest percentage of students with disabilities and ELP students, while having the highest attendance rates and percentage of White students. Oppositely, Schools 09 and 12 had higher rates for students with disabilities, ELP students, and students that qualified for FARMs.

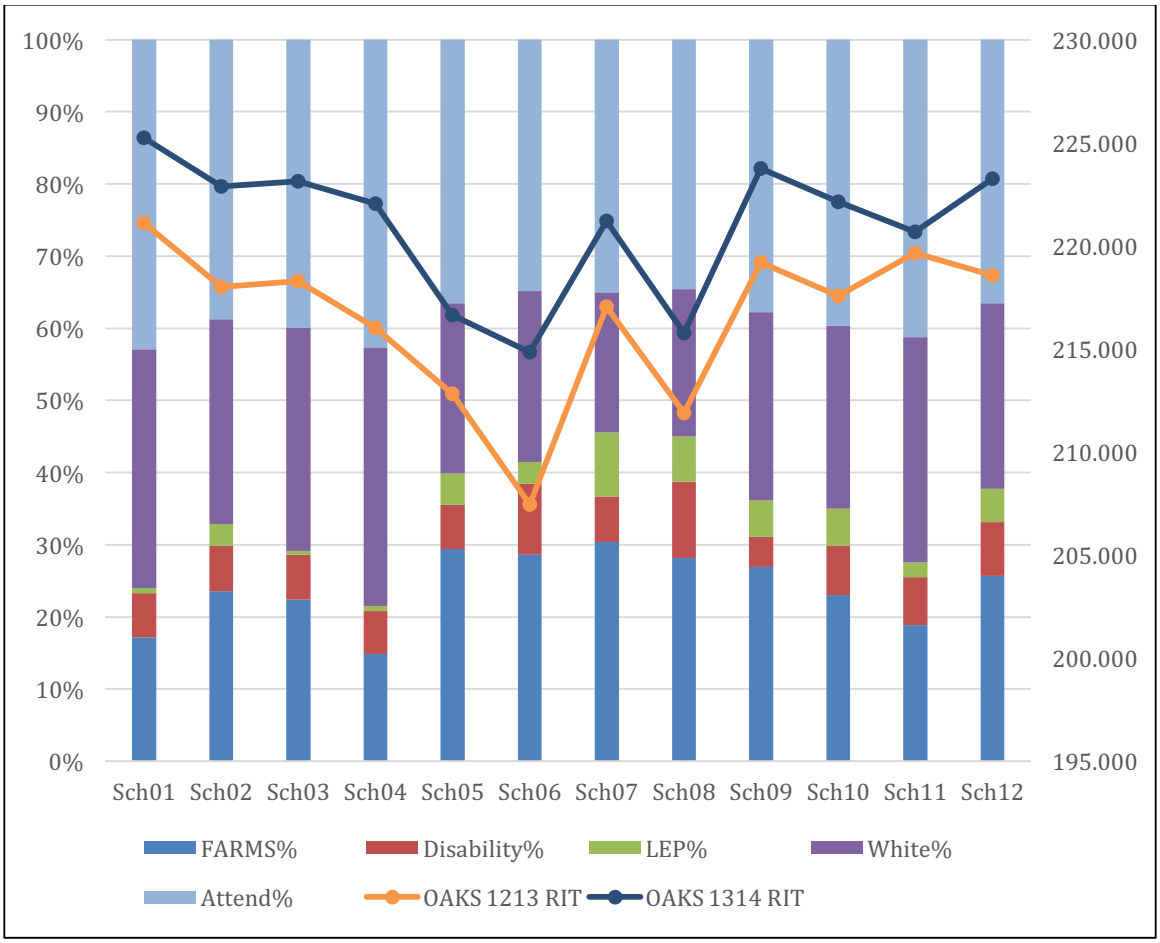

Figure 22. Schools ranked by Summative Principal Ratings and non-academic factors and OAKS-R (stacked).

Results Summary

This study examined the use of fourth and fifth grade summative test score data from the 2013-2014 OAKS-R Reading, and Grade 4 and 5, fall and spring, easyCBM reading curriculum-based, formative test score data (MCRC, PRF, and VOC) aggregated at the building level to assess which assessment type more effectively measures a building principal's impact on student achievement.

Research Question One. The first question analyzed the relationship amongst Summative Principal Ratings (SPR), the spring easyCBM measures (MCRC, PRF, and VOC), and the 2013-2014 OAKS-R Reading statewide test. The results poorly correlated with spring OAKS-R. The highest correlation was between spring OAKS-R and PRF (r = .09), while the lowest correlation was between spring OAKS-R and VOC (r = .05).

Research Question Two. The second question analyzed the relationship between Summative Principal Ratings and easyCBM (a) MCRC, (b) PRF, (c) VOC and (d) OAKS-R with none of these selected measures even moderately predicting Summative Principal Ratings. The quality of the fit for MCRC (R^2 = .02), PRF (R^2 = .36), VOC (R^2 = .07), and OAKS-R (R^2 = .06); meaning that only two percent, 36 percent, seven percent and six percent, respectively, was accounted for by principal ratings.

Research Question Three. The third question analyzed the relationship between Summative Principal Ratings and a Principle Component Analysis (PCA), which grouped demographic risk factors of (a) attendance, (b) ELP, (c) FARMs, (d) percent Other-than-White, and (e) Special Education. When the demographic risk factors were added the quality of fit was low for each factor with ELP (R^2 = .19) being the highest variable, and Special Education (R^2 = 0.00) accounting for zero percent of the Summative Principal

Ratings. When viewing the slope for the line of best fit only attendance had a positive slope ($\beta = .66$), with ELP, FARMs, percent Other-than-White, and Special Education all showing minor negative slopes, depicting slight negative relations.

Research Question Four. The third question visually displayed the academic and non-academic variables. The visual displays offer a graphic rationalization for the poor correlations and lack of statistical significance in Research Questions One, Two, and Three. The displays show that schools with the lowest principal ratings did not have the lowest academic scores.

CHAPTER IV

DISCUSSION

In this chapter, I provide (a) a summary of the results presented in the previous chapter, (b) a review of the limitations to my study, (c) a connection between my results and previous research, (d) a discussion of the practical implications of my results, and (e) suggestions for future research.

Review of Findings

In my study I examined the use of Grade 4 and 5 summative test score data from the spring 2013 and 2014 reading scores from the OAKS-R, and Grade 4 and 5 fall 2013 and spring 2014, easyCBM reading curriculum-based, formative test score data aggregated at the building level to assess which assessment type relates to a building principal's impact on student achievement. The results indicated that Summative Principal Ratings are poor predictors of the academic success of all students whether using large-scale summative assessment (OAKS-R) or formative assessments (easyCBM), with all measures accounting for a fairly small amount of the Summative Principal Rating. However, demographic variables were slightly more related to the Summative Principal Rating.

Substantive Findings: Research Questions One and Two

Results of my study indicated that Principal Summative Rating scores are poorly correlated with the large-scale summative assessment OAKS-R, and poorly correlated with the formative assessment easyCBM. The highest correlation for the formative assessment was the PRF (r = .09) and the OAKS-R correlation was (r = .06). In my study, neither the summative nor formative assessments even moderately predicted the

Summative Principal Ratings. The results also indicated that Summative Principal Ratings are poor predictors of the academic success of all students whether using large-scale summative assessment (OAKS-R) or formative assessments (easyCBM), with both measures statistically accounting for a minor amount of the Summative Principal Rating.

My lack of significant findings and low correlations are not unique to my study; other researchers have found similar results, though the results are mixed. Lipscomb, Chang and Gill (2012), whose study utilized the longitudinal database of all Pennsylvania students in Grades 4-8 between 2007-2008 and 2011-2012, main findings indicated in both math and reading at the elementary and middle school levels that VAM estimates provide little useful information for evaluating elements of principal effectiveness that are persistent over time and common across grades. As a qualifier, Lipscomb et al.'s findings pertained to principals in their first three years at their current position, regardless of length of time as a principal. Lipscomb and colleagues concluded that the differences in principal effectiveness that matter most to evaluators, which are the persistent differences in leadership qualities that impact multiple grades, are not strongly predicated by measures of school value-added.

In contrast to my findings, Grissom, Kalogrides, and Loeb's (2014) Miami-Dade County Public Schools study indicated that VAM model choice could have an important bearing on a principal's effectiveness rating. Likewise, Dhuey and Smith (2014), using data from North Carolina principals at schools serving student in Grades 3-8, found that principal-school match quality accounts for a significant variation in principal value-added. Finally, while supporting positive findings for administrator VAM, Cannon,

Figlio, and Sass' (2012) evidence from Florida, warned that the persistence of a principal's value-added estimates declined when principals changed schools.

When reviewing research on principal VAM measures (incorporating standardized test scores into principal evaluation systems), consideration should also be given to research on teacher VAM measures. Researchers have found that teachers have a large and lasting impact on student achievement (Gordon, Kane, & Staiger, 2006; Rowan, Correnti, & Miller, 2002; Sanders, Wright, & Horn, 1997). While some suggested that VAM scores are the most direct indicators of teacher quality and effectiveness (Gordon, et al., 2006; Hanushek, 2007; Hanushek, Kain, O'Brien, & Rivkin, 2005), other scholars have doubted the accuracy and validity of value-added scores. The doubters (see Amrein-Beardsley, 2008; Kupermintz, 2003; McCaffrey, Lockwood, Koretz, Louis & Hamilton, 2004) suggested that standardized scores may represent some numeric value that teachers add to teaching, but does not account for measurement error, prior teachers' impact, and bias resulting from the placement of teachers into schools and students into classrooms. Amrein-Beardsley, Kupermintz, and McCaffrey et al.'s critiques suggest that value-added scores may fail to accurately represent teacher quality. Agreeing, Koedel and Betts (2007) found that although teachers may have substantial influence over student outcomes in San Diego Public Schools, variance decomposition also showed only low moderate reliabilities of .57 in mathematics and .46 in reading.

Furthermore, two recent studies have shown that teacher value-added scores often vary considerably by the student assessment form and subtests used to construct them (Lockwood, McCaffrey, Hamilton, Stecher, Le & Martinez, 2007). In their study of a commonly used observational system, Classroom Assessment Scoring System (CLASS),

Pianta, Belsky, Vandergrift, Houts and Morrison (2008) found that students' growth trajectories were only moderately predictive, and concluded that the most consistently significant factor, socio-emotional qualities of interactions, matter somewhat when predicting student academic growth. Other studies, including those that compare administrators' ratings of teachers to value-added outcomes, returned low correlations that were between .20 and .50 (see Jacob & Lefgren, 2005; Kimball, White & Milanowski, 2004; Medley & Coker, 1987). In conclusion, my formative and summative research findings mirror the findings of many other researchers who found low predictive ability of students' standardized academic measures to predict adult metrics of job effectiveness, though further study is likely needed given mixed results.

Substantive Findings: Research Question Three

My third question, analyzed the relationship between Summative Principal Ratings and individual demographics risk indicators, and a Principle Component Analysis (PCA), which grouped these demographic risk factors of (a) Attendance, (b) ELP, (c) FARMs, (d) percent Other-than-White, and (e) Special Education. When the demographic risk factors were added the quality of fit was low for each factor with ELP being the most strongly-related variable ($R^2 = .19$), and Special Education accounting for almost none of the Summative Principal Ratings ($R^2 = 0.00$). When viewing the slope for the line of best fit, only Attendance had a positive slope ($\beta = .66$), with ELP, FARMs, Other-then-White, and the PCA grouped demographic analysis demonstrating a minor negative slope, depicting slight negative relations.

The struggle to determine a principal's true contributions to student outcomes from the influence of other school-related factors is a key challenge to calculating

principal VAM (Lipscomb et al., 2010). When attempting to address the direct academic impact that principals have on student achievement, it is important to also consider non-academic factors beyond the administrator's control that may impact student performance (Lipscomb et al., (2012). Leithwood, Louis, Anderson, and Wahlstrom (2004) argued that available studies indicated that principal actions explained between .25 and .34 of the variation in student performance even when figuring in non-academic (demographic) student characteristics, which (when using .25) leaves 93.75% of the variance explained by other factors not in the model and by error.

Similarly, Franco and Seidel (2014) suggested that student non-educational factors contributed to a school effect that was well outside the average academic schooling effect in some value-added models. Contrary to the expected VAM framework, Franco and Seidel's model for Grade 4 mathematics was weak (R^2 between .02 and .05) for students on free/reduced lunch and/or those with high mobility. Franco and Seidel found negative correlations were even stronger for Grades 7 and 8 math for minority students on their value-added scores (R^2= - .31). Franco and Seidel discovered that certain variables were more consistently related to value-added results than others, for example, percent of economically disadvantaged or free/reduced lunch eligible students in a school. The researchers concluded that there was a stronger correlation between non-academic building-wide poverty and value-added ratings than there was for other academic factors. Importantly, those non-academic factors created negative institutional level or building effects among high poverty levels, traditionally underserved populations of students, and VAM. As I detail in the Discussion section, the grouping of the relationships found with certain variables may indicate that these variables are all proxies

for how the school academic factors interact with socioeconomic and family factors shown to predict student achievement in previous research.

In summary, Creemers and Reezigt's (1996) study concluded that leadership only explained three-to-five percent of the variation in student learning across schools, which was actually about one-quarter of the variation (10-20 percent) explained by school-level non-academic (demographic) variables. Similarly, Hill (1998) stated that non-academic classroom factors explained about one-third of the variation in student achievement.

Substantive Findings: Research Question Four

Schools ranked by Summative Principal Ratings and demographics. Figures 13 and 14 display (by school) (a) FARMs percentage, (b) Special Education percentage, (c) ELP percentage, (d) White percentage, and (e) Attendance percentage. Schools were rank-ordered with School 01 having the principal with the highest summative rating and School 12 having the principal with the lowest summative rating.

If the intention of the district were to provide the strongest leadership to the most highly impacted schools, these data seem to indicate this is not occurring. Based on demographic variable data, the principals with the highest summative rating scores are currently placed in the schools with the least impacted students, and the principals in the mid-range on summative rating scores are in the most highly impacted schools. Thus, the district might reconsider their administrative policy and placement procedures in order to place the strongest leadership with the most demographically at-risk populations. It might behoove the district leadership to also analyze their teaching staff placement to see if the similar trends apply.

Schools ranked by Summative Principal Ratings and PRF. Figures 14 and 15 illustrate the growth of students in PRF on the easyCBM formative assessment from fall to spring of 2013-2014. Again, the schools were rank-ordered with School 01 having the principal with the highest summative rating, and School 12 having the principal with the lowest summative rating. The Figure 14 shows school 01, 02, 08, and 09 students scored the highest, with the students in schools 05 and 06 scoring the lowest.

It is notable that the students attending the schools with principals who have the lowest summative rating scores did not have the lowest PRF scores. When reviewing data from this chart the district might want to examine the practices of the teachers in the schools with the highest PRF scores in order to replicate those practices in the lower performing schools.

Schools ranked by Summative Principal Ratings and MCRC. Figures 17 and 18 display the growth of students in MCRC scores on the easyCBM formative assessment from fall to spring of 2013-2014. Again, schools were rank-ordered by the principal's summative rating. As a reminder, students in School 04 had the highest MCRC scores and the students in Schools 07 and 12 showing the most growth between the fall and spring tests. It also shows that the students attending Schools 05 and 11 having minimal growth. When reviewing the data from this chart, the district might consider examining the instructional practices and behaviors of the teachers at schools 07 and 12 in order to replicate those practices in schools where reading comprehension scores showed too little growth.

Schools ranked by Summative Principal Ratings and VOC. Figure 19 and 20 showed the growth of students in VOC scores on the easyCBM formative assessment

from fall to spring of 2013-2014. Again, the schools were rank-ordered by the principal's summative rating, with School 12 having the principal with the lowest summative rating. These data indicate that the students of Schools 02, 09, 12, and 01, in that order, scored the highest. These data indicate that the principal with the lowest summative rating had the students who demonstrated the most growth for vocabulary from fall to spring. Likewise, the principal with the highest summative rating had the students who demonstrated some of the least amount of growth. When reviewing the data from this chart, the district may want to examine the instructional practices and behaviors of the teachers at school 12 in order to replicate those practices in schools where vocabulary scores showed too little growth. Such curriculum and instruction practices could become the focus of teacher-to-teacher district-wide professional development.

Schools ranked by Summative Principal Ratings and OAKS-R. Figures 21 and 22 present the growth of students in the large-scale summative assessment OAKS-R scores from the spring of 2013 to the spring of 2014. Like previously, the schools are rank-ordered with School 01 having the principal with the highest summative rating, and School 12 having the principal with the lowest summative rating. These data indicate that the students in schools 01, 09, and 12 have the highest OAKS-R scores. One factor to consider when analyzing this OAKS-R summative assessment data is while school principals may have remained the same over the 12-13 and 13-14 school years, the students had two different classroom teachers during those same two years. Figure 17 also graphically demonstrates the inherent problem of the district's administrative evaluation tool. If summative gains are an integral part of the administrative evaluation system, then the district might need to re-evaluate their administrative assessment. I

would also suggest that the district look at their teacher evaluation system to see if it suffers the same large-scale test anomalies.

Measuring School Leaders

The idea that effective school leadership is an important characteristic of successful schools has a long history in qualitative studies of effective schools (Purkey & Smith, 1983). For more than 50 years, various presidents have enacted accountability systems through legislation with the intention of establishing standards to hold states, schools, and educators accountable for the academic results of students. Lipscomb, Teh, Gill, Chiang and Owens' (2012) suggested an increasingly common approach across this last decade was to assess principal effectiveness through the use of VAM, which were designed to estimate educators' contributions to student achievement. While some recent quantitative evidence have indicated that more experienced principals have higher school performance (Clark, Martorell, & Rockoff, 2009), a greater ability to recruit and retain high-quality teachers (Branch, Hanushek, & Rivkin, 2012), and greater organizational management skills (Grissom and Loeb, 2011), policymakers continue to be most interested in using student outcomes to measure principal effectiveness for the purposes of accountability directly related to student achievement.

In 2012 when the U.S. Department of Education offered states the opportunity to apply for flexibility from certain aspects of No Child Left Behind, one of the waiver's four core principles was supporting effective instruction and leadership. States were required to: (a) implement VAM evaluation and support systems to improve practice and increase student achievement (U.S Department of Education, 2013), and (b) incorporate student learning and growth as a significant factor in every administrator's summative

performance evaluation. In 2010, Lipscomb, Teh, Gill, Chiang and Owens (2012) described a typical VAM as including an outcome measure, a baseline (or prior year) measurement of the outcome, control variables, and teacher or principal variables. Franco and Seidel (2014) stated that there was not a universal definition for the term value-added. Instead, they suggested the term was used by states to refer to an alternative way to provide feedback of educational effectiveness based on analysis of individual student growth over time.

Oregon's approval of the No Child Left Behind waiver led to the passage of SB 290 (Or. 2011). Guidelines aligned to state and federal legislation required all Oregon districts to incorporate student learning and growth as a significant factor in the overall performance rating of teachers and administrators, and to use student test scores aggregated at the building level through a VAM. While Oregon does not support the use of standardized data as the sole measure of student learning, or the measures of student growth as the sole component on which to base an evaluation, both of these components are a requirement of every building administrator's summative performance evaluation (Oregon Department of Education, 2013).

Limitations

Though my study has several practical implications for practice, it is also subject to several limitations, which must be considered when interpreting these results. The limitations in this study were related to: (a) internal validity, (b) external validity, and (c) construct validity.

Internal validity. Three main internal validity issues affected my study. Those three areas were: (a) sampling, (b) evaluator, and (c) instrumentation.

Convenience sample. The first potential threat to internal validity in this study was my sampling plan (Parker, 1990). All students, elementary principals and the principal evaluator for this study were samples of convenience. Furthermore, all administrative participants were from the same school district in Oregon. Additionally, only those principals were rated on their school's students who took all four measures were included in the analysis. In order for a student to participate in this study, s/he must have met the following criteria: (a) the student was administered the (a) easyCBM PRF, VOC, and MCRC measures in the fall and spring of the 2013-2014 school year, and (b) the computer-based statewide reading/language arts assessment—OAKS-R in the spring of 2012 and 2013.

Principals and evaluator. My study utilized the Summative Principal Ratings of all 12 elementary principals in the district and did not account for length of tenure as an elementary principal or their prior success as a classroom teacher. The Principal Summative Ratings might have been biased by the use of a single individual evaluator of the principals. The single principal evaluator had previously been a principal and colleague to some of the participating principals s/he evaluated. Under such conditions, even if the principals were of equal leadership ability, the principals may receive higher or lower ratings in the areas *professional practice* and *professional responsibilities* based upon factors other than those supposedly assessed by the Principal Rating System.

Instrumentation. The evaluation process and the measurement system used to evaluate principals have been in use for two years. There is no known psychometrics established on the principal rating instrument prior to its implementation. Also, no intra-rater reliability statistics were established for the single principal evaluator, and no inter-rater reliability metrics were evaluated by the district. Thus, the system-level assessment

statistics were not established and the accuracy of the single principal evaluator's observations used in the ratings could affect the dependent (outcome) measure.

External validity. Two main external validity issues affected my study. Those two areas were: (a) population sample, and (b) general of large-scale.

Population sample. This study lacks strong external validity as the result of utilizing a convenience sample of fourth and fifth grade students enrolled in one small district in the Pacific Northwest. The small number of principal participants (n = 12 for principals), a small number of fourth and fifth grade student participants (n = 1,549), and a student population where 68% of students are white. As a contrast, Oregon Department of Education (www.ode.state.or.us) shows the average Hispanic / Latino district make-up as 21.97% and the number of white students to be 64.46% across all Oregon districts. Moreover, the district was 5.3% lower than the average state passing rate for the English Language Arts Assessment for students in grades three through five and 3.1% lower than the average state passing rate for the Mathematics Assessment for students in grades three through five (see district's Oregon Report Card 2014-2015). While close, the lower percentages limit the generalizability to other Oregon districts and school settings with differing administrative and student populations. In order to generalize to other populations, that district administration and student population would have to closely mirror my study's population.

Generalization of large-scale OAKS assessment. The OAKS-R is specific to Oregon. While my study may generalize to other similar districts in Oregon as a result of the requirement that all Oregon districts use the OAKS-R as their summative assessment,

the generalizability of my study to other states is lacking because 49 out of 50 states do not use the OAKS-R.

Construct validity. Two main external validity issues affected my study. Those two areas were: (a) inadequate preoperational explication of constructs, and (b) large-scale assessment aligned to standards.

***Inadequate preoperational explication of constructs*.** All 197 Oregon school districts are required to use evidence of student learning and growth to assist in the evaluation of principal effectiveness according to SB 290. Evidence is collected through multiple sources and principals are scored on a district-designed, four-point, continuous scale rubric. Importantly, each district designed their own evaluation system and specified the types of multiple sources utilized in the administrative assessment and there is lack of commonality across those 197 districts. My study did not assess the evaluation standards, or how the standards are measured from district to district, which limits the conclusions I can draw from my data analyses.

***Large-scale assessment aligned to standards*.** The Oregon Department of Education (2012) reported that the OAKS-R test aligns to Oregon content standards and measures the breadth of the standards. However, the depth and breadth to which teaching to the standards occurs within a district depends on the individual teacher. Differences between what is taught and standard expectations can exist within districts, as well as differences amongst teachers between districts. Again, my study did not assess whether the OAKS-R assessment accurately measures Oregon's curriculum standards for which students are being taught, nor did it assess whether those OAKS-R standards accurately reflected actual teacher instruction.

Future Research

Research on the importance of school leadership and its impact on academic outcomes for students will continue to evolve. My research found that the district's Summative Principal Ratings were poor predictors of the academic success of all students whether using large-scale summative assessment (OAKS-R) or formative assessments (easyCBM). Implications for research and policy-making include the overall need to continue to improve principal evaluation systems, and the study of both the direct and indirect impacts that principals have on student success achievement. When principal evaluation systems are directed by federal and state legislation, the legislation should be predicated upon empirical research findings. If that standard is met, then the federal and state legislation can be a guide in conjunction with other relevant components.

The first relevant component to be further studied is the principal's influence on the school environment and context as a whole. Ladd (2009) found a positive association between student achievement and positive working conditions. Positive working conditions included creating a trusting, collegial, supportive and team-based school culture, promoting ethical behavior, and creating strong lines of communication. Similarly, the Wallace Foundation (2011) stated that working under the supervision of an inspiring and highly competent principal makes the difference in teachers' openness, even eagerness, to work in challenging school environments. Clifford, Hansen and Wraight (2012) found that principals indirectly influence student achievement and instructional quality by creating conditions within schools. Grissom and Loeb (2011) suggested principals leading schools that demonstrated exceptional organizational management (e.g. managing facilities, budgets, and resources and developing a safe

school environment) have a significant relationship to improving student achievement. Finally, Leithwood, Louis, Wahlstrom, and Anderson (2010) concluded from their meta-analysis of principal effectiveness studies that principals' influence student achievement by influencing school contexts.

From a district perspective, there are multiple considerations beyond test scores that districts could use for obtaining data relative to supporting principals through the evaluation system that would link principal evaluation to principal growth. One such consideration is the development of an evaluation system specific to probationary administrators. The district could design tiered rubrics, aligned to ISSLC standards and district goals, specific to probationary years one, two, and three, with the expectations and outcomes clearly defined. Baseline data would be collected in areas to include student attendance, discipline, positive behavior support, and parental involvement, with goals identified based on current data. Additionally, a mentor would be assigned to provide ongoing support, observation and feedback specific to the rubrics and evaluation goals throughout the school year. Ongoing professional development would occur based on district needs and the specific needs of each principal. When considering academic factors, a significant area for growth and evaluation should be on the principals' abilities to establish and maintain a system of ongoing professional development that improves teacher practice within the school setting, and provides every teacher with the support, observation and feedback that increases the teachers' direct impact on student growth and achievement. Future research should include determining what principal factors directly impact teachers' abilities to influence student achievement.

A second component to be considered for future research is the study of highly effective principals, including their experiences as classroom teachers. Given the positive emphasis school leaders place on their prior experiences as teachers (Fiore & Curtin, 1997), it is reasonable to suggest the quality of their experiences play a role in shaping the effectiveness of future leadership behaviors. Fiore and Curtin also stated the research on school leadership suggests that there are specific skills, knowledge, and behaviors that differentiate highly effective principals from those who are less effective. Studies should examine the characteristics, experiences and behaviors utilized throughout their teaching experiences, including behaviors around instructional matters. Summative evaluation scores that occurred during their teaching tenure, as well as both formative and summative student test-score data should be examined.

Finally, teacher quality within schools, and the length of time a teacher and principal have worked together in the same school should likely be a consideration in future studies. According to Public Agenda's (2009) Retaining Teacher Talent study, teachers viewed principal quality as a strong factor in their choice to join or a leave a school. Moreover, Milanowski, Longwell-Grice, Saffold, Jones, Schomisch & Odden (2009) similarly found that principal quality was the most important factor in attracting prospective teachers. What is not addressed is the quality of teachers that are inherited by a new principal and the impacts those teachers have on student achievement, which ultimately has the ability to impact principal quality (and likely their evaluations).

Conclusions

Holding schools to higher standards, closing achievement gaps, increasing equity, and improving the quality of instruction is vitally important to improving educational

outcomes for all students. NCLB's goal of 100% of students reaching the proficient level or higher, while relying on assessment and accountability as a mechanism for improving student achievement, was a worthwhile and important goal, but a goal that proved to be unattainable (Linn, Baker and Betebenner, 2002).

The Obama administration's Flexibility Waiver (U.S. Department of Education, 2012) provided states the opportunity to request relief from certain provisions of No Child Left Behind Act (NCLB, 2001). One of the four core principles was supporting effective instruction and leadership. The waiver required states to implement teacher and principal evaluation systems that provided feedback and support needed to improve practices and increase student achievement through the use of student test scores. In addition, the waiver challenged states to have an effective principal in every school, and an effective teacher in every classroom (U.S. Department of Education, 2010a). States were called upon to evaluate teachers and principals based on student growth and other factors, and systems of evaluation are expected to inform professional development to assist in strengthening the teachers and principals in the schools.

My research examined one district's attempt to implement a principal evaluation system required by Oregon's NCLB waiver and Oregon's Senate Bill 290. In my research, neither the summative or formative assessment informed the district's principal rating system. Specifically, I looked at the required inclusion of school test scores in the yearly evaluation of school principals within the current school reform and the standards and accountability movements. I further narrowed my analysis to fourth and fifth grade fall and spring reading curriculum-based measure scores (easyCBM), and spring-to-spring reading scores from the state's summative accountability assessment (OAKS-R).

As noted earlier, results of my study indicated that Principal Summative Rating scores were poorly correlated with the large-scale summative assessment OAKS-R, and poorly correlated with the easyCBM formative assessments. In my study, neither the summative nor formative assessments even moderately predicted the Summative Principal Ratings. Instead, the results indicated that Summative Principal Ratings are poor predictors of academic success for all students, with all measures accounting for quite a small amount of the Summative Principal Ratings variance. Evidence from my study suggests that administrator summative performance evaluations are not strong indicators of student success, and the ability to adequately assess effective leadership requires more evidence than student test scores. Thus, from a policy implementation perspective, the requirement to incorporate student learning and growth as a significant factor in the overall performance rating of administrators, and to use student test scores aggregated at the building level through a value-added process needs to be carefully considered, and perhaps even reconfigured at the state and individual district levels.

APPENDIX A

DISTRICT ADMINISTRATOR GROWTH AND EVALUATION SYSTEM

Building: **Responsible:**

ADMINISTRATOR SELF-REFLECTION RUBRIC

Standard 1: Visionary Leadership

An educational leader promotes the success of all students by facilitating the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by all stakeholders.

Guiding Questions

- Does the educational leader establish high and measurable expectations for all students and educators?
- Does the educational leader facilitate a process of creating a shared vision, mission, and goals?
- Does the educational leader sustain strong organizational commitment to vision, mission, and goals aimed at continuous improvement?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.1 Believes that all students can learn and incorporates diverse perspectives and builds consensus to provide equitable, appropriate, and effective learning opportunities for every student to achieve.	Informs stakeholder groups about vision, mission, and goals, which are limited in scope and does not emphasize the belief that children can learn.	Invites input from many stakeholder groups in the school community when creating vision, mission, and goals with the belief that most students can learn.	**...and** Includes various stakeholder groups - including the school population and community - and builds consensus about vision, mission, and goals to provide equitable, appropriate, and effective learning opportunities for every student to learn and achieve.	**...and** Provides training, facilitation, and guidance about demographics, research, and best practices to all stakeholder groups and engages them in setting vision, mission, and goals in order to provide equitable, appropriate, and effective learning opportunities in which every student learns and achieves.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.2 Uses varied sources of evidence and analyzes data about current practices and outcomes to shape vision, mission, and goals with high, measurable expectations for all students and educators.	Uses a single source of information and data about what is currently occurring to shape mission and goals.	Uses several information sources and some data to shape vision, mission, and goals with expectations for students and educators.	**...and** Selects and organizes data about current processes and outcomes and connects that data with best information sources that shape vision, mission, and goals with high measurable expectations for all students and educators.	**...and** Engages all stakeholder groups in understanding information from a variety of sources and in understanding how to analyze data in order to assist in focusing vision, mission, and goals with high, measurable expectations for all	N/A

-- DRAFT --

Page 1 of 17

				students and educators.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.3 Aligns, revises, and adapts vision, mission, and goals to school, district, state, and federal policies.	Treats expressions of vision, mission, and goals as isolated statements and programs with little reference to overall policy direction.	Scrutinizes vision, mission, and goals for congruence with school, district, state, and federal policies.	**...and** Ensures that the overall meaning and expressions contained within vision, mission, and goals are aligned with school, district, state, and federal policies.	**...and** Leads stakeholder groups to know and use school, district, state, and federal policies as a method of aligning, adapting, and revising shared vision, mission, and goals within the larger context of school improvement and student achievement.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.4 Communicates and acts on commitments in the shared vision, mission, and goals so educators and the community understand, support, and act on them consistently.	Shows little regard for vision, mission, and goals in reference to actions and decisions.	Supports through words and actions commitment to vision, mission, and goals.	**...and** Leads and models to staff and community the responsibility to act with consistency and adhere through words and actions a commitment to vision, mission, and goals.	**...and** Leads, models, and assesses with staff and community the effectiveness of their adherence to words and actions which demonstrate commitment to vision, mission, and goals and works to ensure continuing fit with school, student, and staff needs.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.5 Uses or develops data systems and other sources of information to identify unique strengths and needs of students, gaps between current	Understands sources of information and data systems, but only uses these sources in limited ways.	Uses information sources and data systems which provide knowledge that can be used to assess whether goals for students are being met.	**...and** Ensures that sources of information and data systems are used to assess whether goals for student progress are being met, and	**...and** Creates or adapts data systems and uses appropriate data systems and additional information sources for assessing goals	N/A

outcomes and goals, and areas for improvement.			then specifies areas for improving student achievement.	and student progress, and proceeds with a plan for involving stakeholder groups in improving student achievement.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.6 Engages faculty, staff and school community by making decisions informed by vision, mission, data, research, and best practices to identify and eliminate barriers, shape plans, programs, and activities, and regularly review their effects.	Has limited knowledge and skill in the use of data and research to identify barriers to goal accomplishment and program review.	Explores and practices with the use of data, research, and best practices to shape plans, programs, and activities associated with vision, mission, and goals.	**...and** Uses collaborative methods with stakeholder groups to review data, research, and best practices in order to shape needed programs and activities, and identify barriers to goal accomplishment by reviewing the impact of the programs and activities on student learning.	**...and** Engages all staff and diverse stakeholder groups in the decision making process and uses data, research, and best practices to identify and eliminate barriers to accomplishing goals by regularly reviewing and evaluating the outcomes of planning, program development, and activities associated with vision, mission, and goals.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
1.7 Uses vision, mission, and goals to guide school improvement planning, change strategies, and implement instructional programs.	Does very little to provide for improvement planning, change strategies, and instructional programs as they relate to vision, mission, and goals.	Consults and considers vision, mission, and goals when planning, selecting change strategies, and selecting instructional programs.	**...and** Aligns vision, mission, and goals with improvement planning processes, change strategies, and implementing effective instructional programs.	**...and** Uses vision, mission, and goals on a regular basis as a foundation for setting short and long-range goals and purposeful alignment of resources to improve instruction.	N/A

Standard 2: Teaching and Learning

An educational leader promotes the success of every student by advocating, nurturing and sustaining a school focused on teaching and learning conducive to student, faculty, and staff growth.

Guiding Questions

- Does the educational leader ensure strong professional cultures that support teaching and learning?
- Does the educational leader establish clear goals and keep those goals in the forefront of the school's attention?
- Does the educational leader require all educators to know and use Oregon Core Curriculum, and current Oregon Effective Teaching Standards (INTASC)?
- Does the educational leader implement appropriate systems of assessment and accountability?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.1 Develops shared understanding of and commitment to ongoing professional learning to achieve high expectations for all students.	Provides professional development with little focus on school needs and with minimal impact on student learning.	Provides sustainable professional learning activities that support student learning and school culture based on school needs.	**...and** Provides sustainable professional learning activities that support student learning and school culture based on school needs.	**...and** Uses multiple data sources to provide varied, targeted and sustainable professional learning activities that enhance adult capacity to positively impact student learning and school culture.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.2 Builds organizational capacity that collectively improves instructional practices and student outcomes.	Makes organizational and instructional decisions with little regard to impact on student outcomes.	Is aware of the impact that hiring, training, and retaining professional staff has on improving instructional practices and student outcomes.	**...and** Consistently hires, trains and retains professional staff that collaboratively work to improve instructional practices and student outcomes.	**...and** Builds a self-sustaining structure that provides for staff to actively participate in improving instructional practices and student outcomes by participating in hiring, training, and instructional decisions.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.3 Builds a professional culture of trust, openness and collaboration, engaging teachers in sharing information, analyzing outcomes, and planning improvement.	Makes school decisions mostly in isolation and informs others of these decisions.	Frequently makes effort to include others in sharing information, analyzing outcomes and planning improvement, but results do not permeate the school culture.	**...and** Builds a professional culture of trust, openness and collaboration, engaging teachers in sharing information, analyzing outcomes, and	**...and** Sustains an ongoing professional culture of trust, openness and collaboration, engaging teachers in sharing information, analyzing outcomes, and	N/A

			planning improvement.	planning improvement.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.4 Uses data about teacher practice to identify and support individual and team professional learning needs across a continuum of professional growth.	Provides very little data-based feedback.	Uses some clearly defined sources of information as the basis for feedback, and offers feedback on an unscheduled, irregular basis.	**...and** Provides feedback that includes multiple clearly defined sources, delivers feedback on a regular basis, and links data and feedback to improvement in teaching and learning for individuals and teams.	**...and** Differentiates some of the data sources on which feedback is based according to individual needs and team needs, and aggregates data to offer individual, school, and system- level feedback which supports a continuum of professional growth.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.5 Guides and supports quality professional development that results in improved teaching and learning and meets diverse learning needs for all students.	Offers little or ineffective professional development without regard to learners' needs.	Initiates professional learning opportunities that apply best practices to classrooms, schools, individual students, and learning objectives for all curricular areas in the aggregate.	**...and** Implements professional learning with performance-based components and ensures that professional learning for educators focuses on the characteristics and demographics of learners.	**...and** Organizes a long-term plan for continuous improvement of quality professional development that is built on data, feedback and continuous study of evidence-based practices and focused on the learning needs for all students.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.6 Works with individual teachers and teams to analyze student work, monitor student progress, adapt curriculum, and adjust instructional practice to meet	Lacks a practical understanding of how to analyze student work, monitor student progress, adapt curriculum and adapt instructional practice.	Knows ways to analyze student work, monitor student progress, adapt curriculum and adapt instructional practice, but does not break it down to individual or team levels.	**...and** Organizes staff and information sources to create a common focus for individuals and teams to analyzing student work, monitoring student progress, adapting	**...and** Uses examples of student and teacher work to clarify and promote individual and team understanding and support collaborative practices that	N/A

student learning needs.			curriculum, and adjusting instructional practice to meet student learning needs.	encourage the use of rigorous curriculum and appropriate instructional practices to promote high levels of achievement for all students.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.7 Facilitates shared understanding of Oregon Core Curriculum and standards-based instructional practices and programs by providing support and guidance to teachers in implementing and aligning rigorous curriculum, effective instruction, content standards, and appropriate assessments to meet student learning needs.	Provides little leadership in applying the Oregon Core Standards, instructional practices, and assessments.	Offers opportunities through random events for educators to understand, apply, and reflect upon the use of Oregon Core Curriculum Standards, instructional practices, and assessments.	**...and** Arranges for professional learning opportunities for staff to understand and apply the Oregon Core Curriculum Standards and reflect upon curriculum, instruction, and assessments.	**...and** Leads staff in formal and informal sessions which focus on Oregon Core Curriculum Standards, aligning rigorous curriculum, effective instruction, and appropriate assessments to meet every student's learning needs.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.8 Provides instructional leadership by supervising, evaluating, coaching, and mentoring educators to improve instructional practice.	Completes supervisory function and encourages professional growth with little success.	Completes formative evaluation tasks and attempts to guide improvement in instructional practice.	**...and** Supervises, evaluates, coaches and mentors educators to improve instructional practice.	**...and** Builds capacity for all educators to be involved in reflective practice to monitor their own progress in improving instruction and assist others in their progress.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*

2.9 Uses effective data-based technology to monitor and promote regular analysis and interpretation of data for accountability to guide continuous improvement.	Uses little data-based technology to direct school improvement.	Studies the use of data-based technology and is learning about the relationship between the technology and school improvement.	**…and** Applies appropriate data-based technology to address goals and to adjust programs and resources for improvement aimed at enhanced teaching and learning.	**...and** Explores cutting edge data-based technologies and other assessment tools, and collaborates with colleagues to improve the quality of those tools for continuous school improvement associated with enhanced teaching and learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.10 Synthesizes data and evaluates progress toward vision, mission, and goals for all stakeholders.	Has little understanding about the relationship between vision, mission, and goals and the use of data.	Studies tools available for gathering data and attempts to relate these data to vision, mission, and goals.	**…and** Organizes and interprets data so that individual educators and stakeholder groups can understand the results and plan for future improvement of vision, mission, and goals.	**...and** Leads educators to organize data so that results have meaning in terms of expressed goals and leads educators to know strategies for communicating complex findings so that individuals and stakeholder groups can understand the meaning of the data in terms of expressed vision, mission, and goals.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
2.11 Planning and developing Student Learning Objectives	Neglects to provide opportunities for teacher collaboration; does not meet with the teacher to look at baseline data, and does not monitor the teacher progress toward the goal.	Works with teacher only occasionally to identify standards or skills to be assessed and may check on measured progress toward the goal.	**…and** Supports the planning and development of SLOs by: Organizes and leads opportunities for collaboration within departments and across grades in developing SLOs; Collaborates with	**...and** Fulfills the requirements for Level 3 and additionally: Utilizes SLOs as the basis of school-wide goals, and vision; Communicates with all stakeholders the purpose and	N/A

			teachers to identify standards or skills to be assessed; Collaborates with teachers to develop/select assessments to evaluate overall student progress. Systematically works with teachers to monitor and revisit SLOs throughout the year; Ensures teachers utilize a tracking tool to show student progress toward SLOs.	progress toward SLOs; Empowers teachers, staff and students to participate in the monitoring of progress toward SLOs and revisits the use and design of teacher and school-wide tracking instruments.	

Standard 3: Management for Learning

An educational leader promotes the success of every student by ensuring management of the organization, operation, and resources for a safe, efficient, and effective learning environment.

Guiding Questions

- Does the educational leader distribute responsibilities and supervise ongoing management structures and practices to enhance teaching and learning?
- Does the educational leader strategically allocate and align human, fiscal, technological, and physical resources?
- Does the educational leader protect the well-being and safety of students and staff?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.1 Makes decisions based on the vision and mission using facts and data. They use a transparent process for making decisions and articulate who makes which decisions. The leader uses the process to empower others and distribute leadership when appropriate.	Provides very few management structures and procedures and is unaware of how they support student learning.	Knows some of the strategies needed to build management structures and procedures and uses them appropriately to support teaching and learning.	**...and** Assigns and reviews the duties and responsibilities associated with leading operational systems and encourages others to assist in the use of appropriate strategies for school management to support teaching and learning.	**...and** Engages stakeholder groups in analyzing duties and responsibilities associated with leading operational systems and planning for appropriate assignments while modeling effective problem solving skills and knowledge of strategic, long-range and operational planning to support teaching and learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.2 Maintains the school facility for safety, ADA requirements, and other access issues to support the learning of every student.	Does not maintain the school facility or keep in compliance with ADA or access issues.	Knows the requirements of maintaining the school facility and works to ensure adherence to safety, ADA, and other access issues.	**...and** Maintains the school facility, adheres to safety and ADA requirements and other access issues in order to support student learning.	**...and** Leads staff and stakeholder groups to understand and sustain the positive relationship between a well-maintained and accessible school facility and student learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.3 Develops and facilitates communication and data systems that assure the timely flow of information to support teaching and learning.	Does little to facilitate the timely flow of information.	Employs systems to provide timely flow of information.	**...and** Develops and facilitates communication and data systems that assure the timely flow of communication to support teaching and learning.	**...and** Engages all diverse stakeholder groups in developing and facilitating communication and data systems to support teaching and learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.4 Operates within budget and fiscal guidelines directed effectively toward teaching and learning.	Has little understanding of monetary guidelines and/or fails to direct resources toward teaching and learning.	Understands monetary guidelines and develops capability of directing resources toward teaching and learning.	**...and** Adheres to monetary guidelines and effectively directs resources toward teaching and learning.	**...and** Enables stakeholder groups to understand the importance of abiding by monetary guidelines and increases their knowledge of ways to effectively direct resources toward teaching and learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.5 Implements practices to recruit, develop, and retain highly qualified and	Does not effectively recruit, develop, and/or retain personnel.	Implements, investigates, and experiments with ways to recruit, develop, and retain highly qualified and	**...and** Recruits, develops, and retains highly qualified and	**...and** Involves other appropriate personnel and diverse	N/A

effective personnel.		effective personnel.	effective personnel.	stakeholder groups as well as teacher leaders to recruit, develop, and retain highly qualified and effective personnel.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.6 Conducts personnel evaluation processes that develop and improve professional practice, by providing timely feedback, and remediates and/or terminates employment in keeping with district and state policies.	Conducts employee evaluations with little regard to adhering to deadlines and procedures.	Employs personnel evaluation processes that offer some benefit in promoting future professional practice and complies with district and state policies.	**...and** Conducts personnel evaluation processes that lead to improved professional practice and complies with district and state policies.	**...and** Relates personnel evaluation policies to improve professional growth and enhance professional practice and complies with district and state polices in ways that all employees can understand.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.7 Advocates for and creates collaborative systems and shared leadership responsibilities that support student and staff safety and well-being.	Does little to create collaborative systems or shared leadership that support student and staff safety and well-being.	Supervises shared responsibilities and collaborative systems that support student and staff safety and well-being.	**...and** Takes responsibility to develop and establish collaborative systems and shared leadership responsibilities that support student and staff safety and well-being.	**...and** Leads others and models an example of how shared leadership and collaborative systems can and do support student and staff safety and well-being.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
3.8 Develops and monitors a comprehensive safety and security plan and implements effective instructional program and practices to	Provides safety and security plan based on insufficient data and investigation and/or may not use the plan as indicated.	Drafts and uses a comprehensive safety and security plan based on critical incidents and needs.	**...and** Develops, shares, and monitors a comprehensive safety and security plan and implements effective instructional	**...and** Collaborates with stakeholder groups and demonstrates to others how the comprehensive safety and security plan relates to a safe environment	N/A

address current safety and well-being topics.	programs and safety practices to address safety topics.	for teaching and learning.	

Standard 4: Community Collaboration

An educational leader promotes the success of every student by collaborating with faculty, staff, parents, and community members, responding to diverse community interests and needs and mobilizing community resources.

Guiding Questions

- Does the educational leader collaborate with families and diverse community stakeholders to extend programs, services, and outreach?
- Does the educational leader respond and contribute to community interests and needs?
- Does the educational leader maximize shared school and community resources to provide essential services for students and families?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
4.1 Collects, analyzes, and shares data and information with the community pertinent to the educational environment.	Makes minimal effort to ensure that public information strategies fit with the abilities and needs of the families and the community.	Experiments with techniques for providing public information strategies appropriate to the needs and abilities of stakeholders in order to communicate with families and community members.	**...and** Collects, analyzes, and shares data and information with the community and uses public information strategies that are appropriate to the needs and abilities of stakeholders in order to communicate with families and community members.	**...and** Shares the rationale behind the data that are collected, analyzed and reported to community members, reviews the efficiency and effectiveness of strategies used to communicate with families and community members, and seeks new ways to continue effective public information strategies that are appropriate to the needs and abilities of stakeholders in order to communicate with families and community members as demographics and opportunities change.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
4.2 Uses effective communication strategies with	Minimally uses communication strategies fit	Experiments with for providing timely and accurate	**...and** Uses effective,	**...and** Shares the	N/A

families and community members to provide stakeholders with timely and accurate information.	without regard to the abilities and needs of the families and the community.	information appropriate to the needs and abilities of stakeholders.	timely, and accurate communication strategies with families and community members appropriate to their needs and abilities.	rationale behind the effective public information strategies chosen, reviews the efficiency and effectiveness of communication strategies with families and community members, and seeks new ways to ensure that effective, timely, and accurate information is communicated to meet the needs and abilities of stakeholders as demographics and opportunities change.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
4.3 Applies communication and collaboration strategies to involve families in decision-making about their student's education and develop positive family partnerships.	Excludes families, whether intentionally or not, and does not involve families in the decision making about their student's education.	Creates opportunities for families to have input into their student's education and understands the importance of family partnerships.	**...and** Involves families in decision making about their student's education, and invites families to partner with the school for improved student outcomes.	**...and** Teaches families of diverse student populations effective strategies for making appropriate, well-informed decisions about their student's educational options and involves the families in the decision-making process through positive family partnerships.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
4.4 Utilizes diversity as an asset of the school community to strengthen educational programs.	Does not acknowledge diversity among students as necessarily desirable or positive.	Studies the ways in which aspects of diverse groups can enrich the school community and strengthen educational programs.	**...and** Capitalizes on diversity as an asset of the school community to strengthen	**...and** Demonstrates through examples and personal modeling the advantages diversity holds for	N/A

			educational programs.	strengthening educational programs and the school community.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
4.5 Shares responsibilities with communities to improve teaching and learning.	Does not share with the community responsibilities for improving teaching and learning.	Attempts in small ways to engage communities in improving teaching and learning.	**...and** Effectively shares responsibilities with communities by inviting stakeholder groups into the school to participate in the improvement process.	**...and** Explains and models to others the behaviors, language, and attitudes that will engage the community in participating in the process to improve teaching and learning.	N/A

Standard 5: Ethical Leadership

An educational leader promotes the success of every student by acting with, and ensuring a system of, integrity, fairness, equity, and ethical behavior.

Guiding Questions

- Does the educational leader demonstrate appropriate, ethical, and legal behavior expected by the profession?
- Does the educational leader examine personal assumptions, values, beliefs, cultural competencies, and practices in service of a shared vision, mission, and goals for student learning?
- Does the educational leader perform the work required for high levels of personal effective leadership performance, including acquiring new capacities needed to fulfill responsibilities?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.1 Models personal and professional ethics, integrity, justice, and fairness and expects the same of others.	Is complacent and/ or unconcerned about personal, professional, and stakeholder groups' ethics, integrity, justice, and fairness.	Sets goals and strives for high standards of personal and professional ethics, integrity, justice and fairness in self and in stakeholder groups.	**...and** Models personal and professional ethics, integrity, justice, and fairness and expects the same of stakeholder groups.	**...and** Articulates, models, expects, and helps to define for self and stakeholder groups the highest standards of professional ethics, integrity, justice, and fairness.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.2 Protects the rights, including appropriate confidentiality, of students and staff.	Inconsistently maintains confidentiality and rarely protects the rights of others.	Meets the requirements of the law with regard to rights and confidentiality of students and staff.	**...and** Protects the rights and appropriate confidentiality of students and staff.	**...and** Models, and leads stakeholder groups to establish ways to protect rights	N/A

				and confidentiality in complex situations.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.3 Behaves in a trustworthy manner, using professional influence and authority to enhance education and the common good.	Engenders little trust or inappropriately uses professional influence and authority.	Is trustworthy as an individual in personal and professional ways.	**...and** Behaves in a trustworthy manner, using professional influence and authority to enhance education and the common good.	**...and** Coaches stakeholder groups to follow practices of consistently trustworthy behavior to enhance professional influence for the common good.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.4 Demonstrates respect for the inherent dignity and worth of each individual.	Shows lack of respect to those in the organization and the community.	Applies generally respectful tones of voice and behaviors with those in the organization and the community.	**...and** Demonstrates respect for the inherent dignity and worth of each individual.	**...and** Engages with stakeholder groups in conversations about a variety of verbal, nonverbal, and organizational ways to demonstrate respect for every person.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.5 Uses a variety of strategies to lead others in safely examining deeply held assumptions and beliefs that may conflict with vision, mission, and goals.	Minimally fosters personal reflection of deeply held assumptions and beliefs that may conflict with vision, mission, and goals by stakeholder groups.	Promotes personal reevaluation of deeply held assumptions and beliefs that may conflict with vision, mission, and goals by stakeholder groups.	**...and** Uses a variety of strategies to lead educators in safely examining deeply held assumptions and beliefs that may conflict with vision, mission, and goals.	**...and** Develops others to use a variety of strategies in safely examining deeply held assumptions and beliefs, including those which may conflict with vision, mission, and goals.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.6 Respectfully challenges and works to change	Selectively ignores various assumptions and	Looks for ways to challenge and change	**...and** Respectfully	**...and** Leads in changing	N/A

assumptions, beliefs, and biases that negatively affect students, educational environments, and every student's learning.	beliefs that negatively affect students, educational environments, and every student's learning.	assumptions and beliefs that negatively affect stakeholder groups and every student's learning.	challenges and works to change assumptions and beliefs that negatively affect stakeholder groups and every student's learning.	assumptions, beliefs, and biases to accommodate the diversity among stakeholder groups and every student's learning.	

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.7 Models lifelong learning by continually deepening understanding and practice related to content, standards, assessment, data, teacher support, evaluation, and professional development strategies.	Lacks evidences of systematic professional growth and its relationships to educational practices.	Seeks opportunities to learn about education, especially content, standards, assessment, data, teacher support, evaluation, and professional development strategies, and strives to see the relationship to effective educational practices.	**...and** Models lifelong learning by continually deepening understanding and practice related to content, standards, assessment, data, teacher support, evaluation, and professional development strategies.	**...and** Leads and models for stakeholder groups a commitment to lifelong learning and effective educational practices, and assists others to see the relationship of these practices to improvement in student learning.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
5.8 Sustains personal motivation, optimism, commitment, energy, and health by balancing professional and personal responsibilities and encouraging similar actions for others.	Does not balance personal and professional commitments and opportunities.	Strives to balance personal and professional commitments and opportunities.	**...and** Sustains personal motivation, optimism, commitment, energy, and health by balancing professional and personal responsibilities and encouraging similar actions for others.	**...and** Teaches stakeholder groups about the concepts and practices related to balancing personal and professional commitments that lead to an emotionally, psychologically, and physically healthy lifestyle.	N/A

Standard 6: Socio-Political Context

An educational leader promotes the success of every student by understanding, responding to, and influencing the interrelated systems of political, social, economic, legal, policy, and cultural contexts affecting education.

Guiding Questions

- Does the educational leader demonstrate active participation in the school's local community?
- Does the educational leader contribute to the larger arena of educational policy-making?

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
6.1 Advocates for equity and adequacy in providing for students' and families' educational, physical, emotional, social, cultural, legal, and economic needs, so every student can meet educational expectations and policy goals.	Does not plan or seek equity or appropriate provisions to ensure opportunities for success for every student.	Plans for changes that could lead to equity or appropriate provisions to ensure opportunities for success for every student.	**...and** Advocates for equity and adequacy in providing for students' and families' educational, physical, emotional, social, cultural, legal, and economic needs, to enable every student to meet educational expectations and policy goals.	**...and** Builds capacity in staff and stakeholder groups to advocate for equity and appropriate provisions to ensure every student can have the opportunities needed for success.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
6.2 Operates consistently to uphold federal, state, and local laws, policies, regulations, and statutory requirements in support of learning for all students.	Shows minimal compliance with federal, state, and local mandates.	Abides by federal, state, and local laws, policies, regulations, and statutory requirements to support student learning.	**...and** Upholds federal, state, and local laws, policies, regulations, and statutory requirements in support of learning for every student.	**...and** Builds capacity in stakeholder groups to understand how to uphold federal, state, and local laws, policies, regulations, and statutory requirements in order to offer support of learning for every student.	N/A

	Unsatisfactory	*Developing Proficiency*	*Proficient*	*Distinguished*	*N/A*
6.3 Communicates effectively with stakeholders in the community and in broader political contexts to improve public understanding of federal, state, and local laws, policies,	Does not communicate educational needs effectively.	Develops awareness of education needs in a variety of contexts.	**...and** Communicates effectively with stakeholders in a variety of ways and in a variety of contexts to improve public understanding of federal, state, and local laws, policies,	**...and** Models and shares strategies and language for communicating effectively with stakeholders the needs of education and to improve public understanding of	N/A

regulations, and statutory requirements.	regulations, and statutory requirements.	federal, state, and local laws, policies, regulations, and statutory requirements in a variety of public political contexts.

APPENDIX B

DISTRICT STUDENT LEARNING AND GROWTH GOALS SAMPLE

Building: **Responsible:**

***ADMINISTRATOR STUDENT LEARNING & PROFESSIONAL GOALS**

Sample Student Learning Goals

Administrator SLG please click here

Student Learning Goal #1

Student Learning Goal
- Goal must address growth for all students, not proficiency

Content Standards/Skills
- What content will the SLG target?

Category 1 - Assessment
Oregon's State Assessments

If required, mark the one you are using

OR

Category 2 - Assessment

Other Assessments (Please list):
- Commercially Developed Assessments that include pre- and post-measures
- Locally development assessments that include pre- and post-measures
- Results from proficiency-based assessment systems
- Locally-developed collections of evidence, i.e. portfolios of student work that include multiple types of performance

Context/Students
- Are there any contextual factors that may impact student growth?
- Include number of students, gender, race/ethnicity, socioeconomic status, diverse learners and contact time

Baseline Data
- What are the learning needs of the students?
- Include summary of student strengths and weaknesses based on data analysis

Student Growth Goal (Targets)

Rationale
- Describe how the focus of the goal was determined

Strategies
- Include stategies used by the educator to support meeting the needs for student growth

Professional Learning & Support
- Identify areas of additional learning and support needed to meet the SLG

Student Learning Goal #2

-- DRAFT --

Page 1 of 3

*Administrator Student Learning & Professional Goals -- DRAFT --

Student Learning Goal
- Goal must address growth for all students, not proficiency

Content Standards/Skills
- What content will the SLG target?

Category 1 - Assessment
Oregon's State Assessments

If required, mark the one you are using

OR

Category 2 - Assessment

Other Assessments (Please list):
- Commercially Developed Assessments that include pre- and post-measures
- Locally development assessments that include pre- and post-measures
- Results from proficiency-based assessment systems
- Locally-developed collections of evidence, i.e. portfolios of student work that include multiple types of performance

Context/Students
- Are there any contextual factors that may impact student growth?
- Include number of students, gender, race/ethnicity, socioeconomic status, diverse learners and contact time

Baseline Data
- What are the learning needs of the students?
- Include summary of student strengths and weaknesses based on data analysis

Rationale
- Describe how the focus of the goal was determined

Strategies
- What will you do differently to help students attain this growth?
- What professional development opportunities will best support the student achievement goals set forth in this SLG?
- Include stategies used by the educator to support meeting the needs for student growth

Professional Learning & Support
- Identify areas of additional learning and support needed by the educator to meet the SLG

Professional Growth Goal

Professional Growth Goal

Strategies
- What will you do to attain this goal?
- What professional development opportunities will best support this goal?

Professional Learning & Support
- Identify areas of additional learning and support needed to meet your professional growth goal

REFERENCES CITED

Alonzo, J., Park, B., & Tindal, G. (2008). The Development of Middle School Passage Reading Fluency Measures for use in a Progress Monitoring Assessment System. (Technical Report No. 46) Eugene, OR: Behavioral Research and Teaching, University of Oregon.

American Recovery and Reinvestment Act of 2009. Pub. L. No. 111-5, 123 Stat. 115, 516 (Feb. 19, 2009).

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75. doi: 10.3102/0013189x08316420

Anderman, E. M., Anderman, L. H., Yough, M. S., & Gimbert, B. G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist, 45*(2), 123-137. doi: 10.1080/00461521003703045

Armour-Garb, A. (2009). Should value-added models be used to evaluate teachers? *Journal of Policy Analysis and Management, 28*, 692-693. doi: 10.1002/pam.20462

Beimers, J. N. (2008). *The effects of model choice and subgroup on decision in accountability systems based on student growth.* University of Iowa, UMI Dissertations (3347200).

Borkowski, J. W., & Sneed, M. (2006). Will NCLB improve or harm public education. *Harvard Educational Review, 76*, 503-525.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals (Working Paper No. 17803). Washington, DC: National Bureau of Economic Research.

Cannon, S., Figlio, D., & Sass, T. (2012). Principal quality and the persistence of school policies. (Working Paper). Washington, DC: National Center for Analysis of longitudinal data in Education Research.

Clark, D., Martorell, P., & Rockoff, J. (2009). School principals and school performance. (Working Paper No. 38). Washington, DC: National Center for Analysis of longitudinal data in Education Research.

Clifford, M., Hansen, U. J., & Wraight, S. (2012). A practical guide to designing comprehensive principal evaluation systems: A tool to assist in the development of principal evaluation systems. Washington, DC: National Comprehensive Center for Teacher Quality.

Coggshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jacques, C. (2012). Generating

teacher effectiveness: The role of job-embedded professional learning in teacher evaluation. In National Comprehensive Center for Teacher Quality (Ed.), *Research & Policy Brief*.

Creemers, B. P., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement*,*7*, 197-228.

Dee, T. S., & Jacob, B. A. (2010). The impact of No Child Left Behind on students, teachers, and schools. *Brookings Papers on Economic Activity,* 149-194.

Dhuey, E., & Smith, J. (2014). How school principals influence student learning. (IZA Discussion Paper No. 7949). Bonn, Germany: Institute for the Study of Labor.

Duncan, A. (2009, September). *Reauthorization of ESEA: Why we can't wait.* Speech presented at monthly stakeholders' meeting. Washington, DC.

Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10 (1965).

Fiore, T. A., & Curtin, T. R. (1997). Public and private school principals in the United States: A statistical profile, 1987-88 to 1993-94. Washington, DC: United States Government Printing Office, Superintendent of Documents.

Franco, M. S., & Seidel, K. (2014). Evidence for the Need to More Closely Examine School Effects in Value-Added Modeling and Related Accountability Policies. *Education and Urban Society, 46*(1), 30-58. doi: 10.1177/0013124511432306

Gardner, D. P., Larsen, Y. W., & Baker, W. (2005). A nation at risk: The imperative for educational reform. United States Department of Education: Washington DC.

Geier, B. A. (2014). The federal government's history in public education: Massive reform efforts for political and corporate enhancement. Tampa, FL: University of South Florida. Retrieved from http://works.bepress.com/brett_geier/2/

Goals 2000: Educate America Act, Pub. L. No. 103-227 (1994).

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). Identifying effective teachers using performance on the job. Washington, DC: Brookings Institution.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness how perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*, 1091-1123.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2014). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 0162373714523831.

Haar, J. (2004). The role of professional development in the principalship. *Catalyst for*

Change, 33(2), 20–24.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, *9*, 157-191.

Hanushek, E. A. (2007). The single salary schedule and other issues of teacher pay. *Peabody Journal of Education*, *82*, 574-586.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teacher quality (Working Paper No. 11154). Cambridge, MA: National Bureau of Economic Research

Hill, P. W. (1998). Shaking the foundations: Research driven school reform. *School Effectiveness and School Improvement*, *9*, 419-436.

Hungerford, N., & Dickson, K. (2012). Oregon's Senate Bill 290: A Primer.

Improving America's Schools Act of 1994, Pub. L. 103-382, 20 U.S.C. 6301 et. Seq. [1994].

Interstate School Leaders Licensure Consortium. (2008). Educational Leadership Policy Standards: ISLLC 2008.

Jacob, B. A., & Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education (Working Paper No. 11463). Cambridge, MA: National Bureau of Economic Research.

Jacobson, S. L., & Bezzina, C. (2008). Effects of leadership on student academic/affective achievement. In J. Lumby, G. Crow, & P. Pashiardis (Eds.), *International Handbook on the Preparation and Development of School Leaders* (pp. 81–103). New York: Routledge.

Kanter, H. (1991). Education, social reform, and the state: ESEA and federal education policy in the 1960's. *American Journal of Education, 100*, 47-83.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, *79*(4), 54-78.

Koedel, C., & Betts, J. R. (2007). Re-examining the role of teacher quality in the educational production function (pp. 2007-03). Nashville, TN: National Center on Performance Incentives, Peabody College, Vanderbilt University.

Kucerik, E. (2002). No Child Left Behind Act of 2001: Will it live up to its promise. *Georgetown Journal on Poverty Law & Policy, IX*(2), 479-487.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis*, *25*, 287-298.

Ladd, H. F. (2009). Teachers' perceptions of their working conditions: How predictive of policy-relevant outcomes? (Working Paper No. 33). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Leithwood, K., Louis, K. S., Wahlstrom, K. L., & Anderson, S. E. (2010). Learning from leadership: Investigating the links to improved student learning. Alexandria, VA: Educational Research Service.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How leadership influences student learning. Minneapolis, MN: Center for Applied Research and Educational Improvement, University of Minnesota.

Levine, M., & Levine, A. (2012). Education deformed: No Child Left Behind and the Race to the Top. This almost reads like our business plans. *American Journal of Orthopsychiatry, 82*(1), 104-113. doi:10.1111/j.1939-0025.2011.01142.x

Levine, M., & Levine, A. (2014). Charters and foundations: Are we losing control of our public schools? *Journal of Orthopsychiatry, 84*(1), 1-6. doi:10.1037/h0098942

Linn, R. L., Baker, E. L., & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the no child left behind act of 2001. *Educational Researcher, 31*(6), 3-16.

Lipscomb, S., Chiang, H., & Gill, B. (2012). Value-added estimates for Phase 1 of the Pennsylvania Teacher and Principal Evaluation Pilot. Cambridge, MA: Mathematica Policy Research, Inc.

Lipscomb, S., Teh, B., Gill, B., Chiang, H., & Owens, A. (2010). Teacher and principal value-added: Research findings and implementation practices. Cambridge, MA: Mathematica Policy Research, Inc.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, *44*(1), 47-67.

Manna, P. (2011). *Collision course: Federal education policy meets state and local realities*. Washington DC: CQ Press.

Martin, V., & Lázaro, L. M. (2011). The race to educational reform in the USA: The Race to the Top. *Language and Education, 25*, 479-490. doi: 10.1080/09500782.2011.596280

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67-101.

McClure, P. (2008). The history of educational comparability in Title I of the Elementary and Secondary Education Act of 1965. *Ensuring Equal Opportunity in Public Education* (pp. 9-16): Center for American Progress.

McDonnell, L. M. (2005). No Child Left Behind and the federal role in education: Evolution or revolution? *Peabody Journal of Education, 80*(2), 19-38. doi: 10.1207/S15327930pje8002_2

McDonnell, L., McLaughlin, M., & Morison, P. (1997). Educating one and all: Students with disabilities and standards-based reform. Washington, DC: National Research Council.

McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants, and the Obama education agenda. *Educational Policy, 26*(1), 136-159. doi: 10.1177/0895904811425911

McNeil, M. (2009). Racing for an early edge. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2009/07/09/36stimulus_ep.h28.html

McNeil, M. (2011). States: Stimulus aid sparked progress on goals. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2011/05/25/32sfsf.h30.html

Medley, D. M., & Coker, H. (1987). How valid are principals' judgments of teacher effectiveness? *The Phi Delta Kappan*, *69*(2), 138-140.

Milanowski, A. T., Longwell-Grice, H., Saffold, F., Jones, J., Schomisch, K., & Odden, A. (2009). Recruiting new teachers to urban school districts: What incentives will work? *International Journal of Education Policy and Leadership*, *4*, 1-13.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal, 84*(2), 112-130.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common Core State Standards. Washington, DC: Authors.

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

Olney, M. (2012, September 26). Letter to Lindsey Capps, Oregon Education Association. Retrieved from http://www.oregoned.org/images/pages/Collaborative_Process_Opinion_Letter_9-26-2012.pdfon.

Oregon Department of Education. (2007). *2006-2007 Technical report Oregon's Statewide Assessment System: Reliability and Validity, Volume 4*. Retrieved from http://www.ode.state.or.us/teachlearn/testing/manuals/2007/asmttechmanualvol4_validity.pdf

Oregon Department of Education, Office of Assessment and Information Services. (2012b). *Oregon statewide assessment: Reading/literature test specifications and blueprints, Grade 4 2012-2014.* Retrieved from http://www.ode.state.or.us/wma/teachlearn/testing/dev/testspecs/asmtrdtestspecsg4_2012-2014.pdf

Oregon Department of Education. (2012). *Oregon framework for teacher and administrator evaluation and support*.

Oregon Department of Education. (2013). *Oregon framework for teacher and administrator evaluation and support systems*.

Oregon Department of Education. (2014). *Oregon framework for teacher and administrator evaluation and support systems*.

Oregon School Boards Association. (August 2013). Oregon's No Child Left Behind waiver. Retrieved from http://www.osba.org/Resources/Article/Improving_Education/SAG/NCLBWaiver.aspx

Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*, 163-193. doi: 10.3102/0002831210362589

Parker, R. M. (1990). Power, control and validity in research. *Journal of Learning Disabilities*, *23*, 613–620.

Parylo, O., Zepeda, S. J., & Bengtson, E. (2012). Principals' experiences of being evaluated: A phenomenological study. *Educational Assessment, Evaluation and Accountability, 24*, 215-238. doi: 10.1007/s11092-012-9150-x

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*, 365-397.

Public Agenda. (2009). Retaining teacher talent survey of teachers: Full survey data. New York: Author. Retrieved from http://www.learningpt.org/expertise/educatorquality/genY/FullSurveyData.pdf

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal*, *83*, 427-452.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools. *The Teachers College Record*, *104*, 1525-1567.

S. 252, 2011 Leg., 76th Sess. (Or. 2011).

S. 290, 2011 Leg., 76th Sess. (Or. 2011).

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education*, *11*(1), 57-67.

Smith, T. M., Desimone, L. M., & Ueno, K. (2005). Highly qualified to do what? The relationship between NCLB and teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis, 27*(1), 75-109.

Thomas, J. Y., & Brady, K. P. (2005). The Elementary and Secondary Education Act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education, 29*, 51-67.

U.S. Department of Education. (2008). A nation accountable: Twenty-five years after A Nation at Risk. Washington DC.

U.S. Department of Education. (2010a). ESEA blueprint for reform. Washington DC: Office of Planning, Evaluation and Policy Development.

U.S. Department of Education. (2010b). *Race to the Top application for initial funding* Washington DC: United States Department of Education. Retrieved from www2.ed.gov/programs/racetothetop/ application.doc

U.S. Department of Education. (2012). *ESEA flexibility request*. (1810-0708). Washington DC: United States Department of Education.

U.S. Department of Education. (2013). *ESEA flexibility: Guidance for renewal process For Windows 1 and 2 States*. Washington DC: United States Department of Education.

The Wallace Foundation. (2011). Research findings to support effective educational policies: A guide for policymakers (2nd ed.). New York: Author.

Whilden, B. E. (2010). The Elementary and Secondary Education Act: A primer on reauthorization in 2010.