

January 1st, 2016

The Relationship Between Student Performance on a Reading Progress Monitoring Measure and the Washington State Standardized Test

Miriam Mickelson
Seattle Pacific University

Follow this and additional works at: https://digitalcommons.spu.edu/soe_etd

 Part of the [Curriculum and Instruction Commons](#)

Recommended Citation

Mickelson, Miriam, "The Relationship Between Student Performance on a Reading Progress Monitoring Measure and the Washington State Standardized Test" (2016). *Education Dissertations*. 15.
https://digitalcommons.spu.edu/soe_etd/15

This Dissertation is brought to you for free and open access by the Education, School of at Digital Commons @ SPU. It has been accepted for inclusion in Education Dissertations by an authorized administrator of Digital Commons @ SPU.

The Relationship Between Student Performance on a Reading Progress Monitoring
Measure and the Washington State Standardized Test

by
Miriam M. Mickelson

Dissertation presented to the
Faculty of the Graduate School of Education at
Seattle Pacific University
in partial fulfillment of the requirements
for the degree of
Doctor of Education

Seattle Pacific University

September 2016

The Relationship Between Student Performance on a Reading Progress Monitoring Measure and
the Washington State Standardized Test

by

Miriam M. Mickelson

A dissertation submitted in partial fulfillment

of the requirement of the degree of

Doctor of Education

Seattle Pacific University


2016

Approved by  _____

Dr. Thomas Alsbury, Chairperson of the Dissertation Committee

 _____

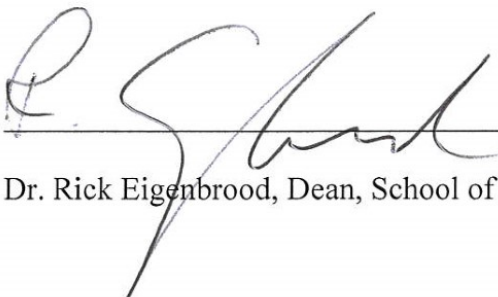
Dr. John Bond, Dissertation Committee Member

 _____

Dr. William Prenevost, Dissertation Committee Member

Program Authorized to Offer Degree SCHOOL OF EDUCATION

Date OCTOBER 2016

 _____
Dr. Rick Eigenbrood, Dean, School of Education

Copyright Page

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at Seattle Pacific University, I agree that the library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the W.S. Copyright Law.

Signature

Miriam M. Michelson

Date

11-13-16

Table of Contents

List of Tables	iv
List of Figures	v
List of Appendices	vi
Acknowledgement	vii
Chapter One: Introduction	2
Purpose of the Study	6
Significance of the Study	6
Research Questions	12
Research Methods	13
Terms and Definitions	14
Chapter Two: Literature Review	17
Theoretical Framework	18
Structural Cognitive Modifiability Theory	18
Empirical Research on Third Grade Reading	22
Empirical Research on the Impact of Reading on Academic Success	26
Empirical Research on Benchmark Measurements	29
Chapter Three: Methods	40
Research Design	40
Participants	41

Instrumentation	42
STAR Reading.....	42
Smarter Balanced Assessment (English Language Arts/Literacy).....	47
Procedure	50
Data Analysis	51
Chapter Four: Results	55
Data Results I.....	56
Data Results II.....	61
Data Results III	74
Summary.....	84
Chapter Five: Discussion	86
Summary of Findings.....	86
Discussion.....	87
Research Question 1	87
Research Question 2	89
Research Question 3	91
Connections to Previous Research.....	93

Reading Achievement Gap	93
Impact of Reading on Academic Achievement.....	95
Reading Progress Monitoring Measures and Standardized Exams.....	96
Limitations of the Study.....	97
Length of Data Collection	97
Missing Data.....	98
Limited School District Participation	99
Recommendations for Future Study	100
References.....	103
Appendices.....	115

List of Tables

Table 1: Washington State and District Demographics	41
Table 2: Third Grade STAR Reading Percentile Rank Range	46
Table 3: Hierarchical Regression Model Summary	67
Table 4: Hierarchical Multiple Regression Coefficients	71
Table 5: Hierarchical Multiple Regression Frequency Statistics.....	73
Table 6: Multiple Regression Model Summary	78
Table 7: Multiple Regression Coefficients	80
Table 8: Multiple Regression Frequency Statistics	84

List of Figures

Figure 1: Hierarchical Multiple Regression Scatterplot Showing Homoscedasticity	65
Figure 2: Multiple Regression Scatterplot Showing Homoscedasticity	76

List of Appendices

Appendix A: STAR Reading Skills and Domains.....	115
Appendix B: Grade 3 SBA English Language Arts/Literacy Targets	116
Appendix C: Hierarchical Regression – Linear Relationship.....	117
Appendix D: Hierarchical Multiple Regression – Normality of Distribution	120
Appendix E: Multiple Regression – Linear Relationship	121
Appendix F: Multiple Regression – Normality of Distribution.....	124
Appendix G: IRB Form	125

Acknowledgements

This dissertation represents the culmination of an educational journey that would not have been possible without the support of so many people. I wish to share this accomplishment with the following.

My loving and supportive husband, David. Thank you for changing my life.

My beloved son, Zach. I have been going to school since your infancy, and now that you are fifteen, I am finally done. Thank you for generously sharing your childhood with a mother who worked, studied, and pursued her dreams.

My parents and my siblings. Thank you for your love, prayers, and encouragement.

The Mickelson's. I am blessed to be a part of your family.

Dr. Alsbury, dissertation chair extraordinaire. I am most grateful for your timely and meaningful feedback, your guidance, and your encouragement. I would not have been able to finish this seemingly insurmountable task without your help.

Dr. Prenevost and Dr. Bond. Thank you for serving on my dissertation committee and for the countless invaluable lessons in educational leadership that you imparted with me since my days taking Principal Certification classes at SPU.

Dr. Mvududu. Thank you very much for patiently answering all my statistics questions.

All my professors at SPU. Thank you for a first-class education in a supportive setting.

Dr. Kimberly Jensen. You have supported me in this journey in so many ways. I am thankful for your friendship.

My work colleagues and friends. I am appreciative of your support and encouragement.

Above all, my Heavenly Father. Thank you for Your steadfast and unwavering love. I do not deserve it, but You freely and unconditionally give it to me nevertheless.

Abstract

The Relationship Between Student Performance on a Reading Progress Monitoring Measure and the Washington State Standardized Test

by

Miriam M. Mickelson

Seattle Pacific University

Dissertation Chair: Dr. Thomas Alsbury

Some experts and educators believe that learning to read is critical to success in and out of school (Lonigan & Phillips, 2015; Nation and Norbury, 2005; O'Connor & Klein, 2004). Schools, therefore, have the responsibility to ensure that all students become proficient readers, especially by the end of third grade, considered to be a pivotal year for literacy (Hernandez, 2011; Lesnick, Goerge, Smithgall, & Gwynne, 2010). Promoting literacy entails providing students with effective literacy instruction shaped and guided by timely, reliable, and meaningful assessment results. Assessment should inform instruction, which hopefully leads to student mastery of state mandated reading standards. This study was conducted to investigate the relationship between student performance on STAR Reading, a progress monitoring measure in third grade, and the Smarter Balanced Assessment in English Language Arts/Literacy, Washington State's standardized test. The relationship between the Smarter Balanced Assessment scores and other variables such as student gender, Free and Reduced Lunch status, and Special Education status was also explored. A multiple methods research design that included Spearman's rank-order correlation, hierarchical multiple regression, and standard multiple regression was utilized to answer the research questions presented in this study.

Findings revealed a statistically significant relationship between STAR Reading and Smarter Balanced Assessment English Language Arts/Literacy scores. Of the three assessment periods for STAR Reading, spring had the strongest statistically significant relationship to the state standardized test compared to the fall and winter test administration periods. Additionally, a statistically significant relationship was measured between the Smarter Balanced Assessment scores and gender, Free and Reduced Lunch status, Special Education status, and STAR Reading scores. Further research is warranted to further explore the relationship between student performance on a reading progress monitoring measure and a state standardized exam.

***Keywords:** interim assessment, benchmark assessment, progress monitoring measure, curriculum-based assessment, standardized assessment*

Chapter One

Introduction

Learning to read is undoubtedly a fundamental skill young students need to master as part of their schooling experience. In primary education, early literacy development, a critical phase in learning and knowledge acquisition, ought to be an elemental goal of educators for every child (Center for Public Education, 2015; Reutzel, 2015). Learning to read during a student's formative years paves the way for reading to learn, which is crucial to success in later academic years and in life (Lonigan & Phillips, 2015). Learning to read ensures one can read in order to learn and access information, which is still primarily delivered in print and text even in today's world of technological advancements. The ability to read and to comprehend what was read helps secure positive learning outcomes, which then leads to a better quality of life as a fully functioning, independent, and contributing member of society (Nation & Norbury, 2005; O'Connor & Klein, 2004). Therefore, it is incumbent upon schools to see to it that all students are reading proficiently early in their educational careers to avert reading and academic challenges later on (Snow, Burns, & Griffin, 1998). This is essential because children with well-developed literacy skills early in their schooling are likely to be proficient readers by the time they leave elementary school. Conversely, children with substandard literacy skills in their primary years are likely to continue having deficiencies in their reading skills as they progress in their education unless they are afforded extensive and appropriate intervention and support (Duncan, et. al, 2007; Good, Simmons, & Kame'enui, 2001; Juel, 1998; Wagner, Torgesen, & Rashotte, 1994; Wagner, et al., 1997). Moreover, struggling readers are alarmingly four times more likely to drop out of

high school and are less likely to earn a decent living wage as adults compared to their peers who are proficient readers (Hernandez, 2011).

Some experts believe students should master literacy by the time they leave third grade lest they face significant academic challenges for the remainder of their educational career (Center for Public Education, 2015; Feister, 2013; Hernandez, 2011; Lesnick, Goerge, Smithgall, & Gwynne, 2010; Miles & Stipek, 2006). Students are well served when their teachers make it their primary goal to help students leave third grade ready to encounter an even wider array of texts in fourth grade, equipped with the skills necessary to engage in analytical reading and to enrich their vocabulary through reading, among other skills (O'Brien, 2008).

Unfortunately, not all third graders move on to fourth grade reading proficiently, and those who do not are disproportionately minority students and students from low income families (National Assessment of Educational Progress, 2011, 2015). For instance, students from lower income families scored 29 points lower than students from higher income families on the 2011 National Assessment of Education Progress (NAEP) reading test, while minority students scored 25 points lower than their peers on the same reading test (National Assessment of Educational Progress, 2011). There is also heightened concern that boys underachieve in reading compared to girls. According to Clark and Burke (2012), girls outperformed boys on all 2012 National Curriculum reading tests in the United Kingdom, while international comparison studies revealed a widening gender gap in reading enjoyment and reading frequency in favor of girls. Results from the National Assessment of Educational Progress (2015) reading test in the state of Washington showed that female students had an average score that was higher

than that of male students by 13 points, outcomes that were consistent with the national trend. Walker (2015) and Loveless (2015) pointed to a continued reading achievement gap between boys and girls, with boys lagging behind their female peers. Another reading achievement gap of note is the gap between students with disabilities and those who are not served in Special Education (SPED) programs. NAEP (2015) reports that in 2015, students without disabilities generally outperformed Special Education students in reading.

Cognizant of the significance of early literacy and various reading achievement gaps, the federal government, through the No Child Left Behind Act of 2001, issued a literacy initiative called Reading First Initiative (Title 1, Part B, Subpart 2). The National Conference of State Legislatures (n.d.) described the Reading First Initiative as a concerted, focused nationwide effort to help all students become proficient early readers by eliminating reading deficiencies through high quality, comprehensive reading programs in kindergarten through third grade. Reading First Initiative calls for teacher professional development on scientifically based and research supported reading instruction and requires accountability of student learning through ongoing, valid, and reliable screening, diagnostic, and classroom-based reading measures.

Screening, diagnostic, and classroom-based reading assessments play a critical role in helping students attain reading proficiency because they provide useful data that form an objective basis for instructional adjustments. Good assessment practices are intimately linked to and should answer important questions about the impact and effectiveness of teacher instruction (Adams, Anderson, & Durkin, 1978; Marzano, Pickering, & McTighe, 2005; McGlinchey & Hixson, 2004; Popham, 2008). Assessments

should therefore be effectively woven into every educator's teaching repertoire. For instance, classroom-based formative assessments, which according to Brookhart (2004) "gives assessment information that is useful for continued student learning, positive classroom change, and other improvements" (p. 6), should be a key factor that informs teacher instruction. Wiliam (2011) posited that because teachers cannot predict what students learn as a result of any particular sequence of instruction, they must adopt sound assessment practices, such as the use of formative assessments, to collect the best possible evidence about student learning and to use that data to decide on next steps.

As a complement to classroom-based formative assessments, schools and school districts are also investing in interim assessments in efforts to gain access to important additional information about student learning. In the area of reading and literacy, it is not uncommon for school districts to utilize and administer progress monitoring measures such as DIBELS, or Dynamic Indicators of Basic Early Literacy Skills, and STAR Reading assessment to gauge students' literacy skills, measure student growth over time, and identify which students are in need of intervention, remediation, enrichment, or further diagnostic examination. Schools invest time, financial resources, and energy on the administration of such tests to predict students' year-end growth, determine students' at-risk status, and foresee the number of students likely to meet state standards at the end of the year as evidenced by state-mandated exams.

Learning to read is an integral part of our students' learning process, and educators must be fully committed to helping every single one of their students achieve reading proficiency. Reading proficiency leads to numerous positive outcomes, while reading deficiencies put students at a severe disadvantage, academically, personally, and

professionally. In order to provide meaningful, relevant literacy instruction as well as appropriate, needs-based, and targeted intervention and enrichment, educators must analyze and take action on data from a number of assessments, including formative and interim tests.

Purpose of the Study

The primary purpose of this study is to examine the relationship between the performance of third grade students on a reading progress monitoring measure and the Washington State standardized test. Specifically, this study seeks to determine if third grade student performance on the STAR Reading assessment has a statistically significant relationship to the Washington State standardized examination called Smarter Balanced Assessment (SBA) in English Language Arts (ELA) and Literacy, which measures student progress towards meeting Common Core State Standards in reading, writing, speaking and listening, and research and inquiry. A secondary purpose of the study is to examine the relationship between SBA ELA/Literacy and STAR Reading, along with other pertinent factors that include gender, Free and Reduced Lunch status, and Special Education status.

Significance of the Study

This study contributes to research and the field of teaching in practical, substantive, and theoretical significance. From the classroom perspective, this study is practically significant in that it helps teachers understand whether performance on STAR Reading translates to proficiency on the state assessment in SBA English Language Arts/Literacy, which measures student progress towards meeting the Common Core State Standards. If STAR Reading is aligned with state standards and correlates with student

performance on the state test measuring students' literacy skills, teachers may feel more committed to administer the test with fidelity and to carefully examine the results in order to make adjustments to their literacy instruction. The results of progress monitoring measures such as STAR Reading may be helpful to teachers not only in predicting literacy and academic outcomes, but also, more importantly, in making appropriate and necessary changes to their instruction in order to promote early literacy of all students, especially those who are at risk of reading challenges. Gambrell, Morrow, Neuman, and Pressley (1999) suggested that effective reading practices are the consequence of informed decision-making. A progress monitoring measure that helps teachers determine students' current reading levels as well as anticipate students' level of mastery of state standards, as measured by the SBA, allows teachers to provide needs-based intervention and enrichment for students. This is especially significant for struggling readers who need individualized and differentiated support that target their learning gaps.

Additionally, when formative and summative assessments are aligned and linked with one another, there is coherence in instruction. Coherent instruction invites teachers to think systematically about their practice, which, according to the National Board for Professional Teaching Standards (1989), contributes to effective teaching. The Danielson Frameworks for Teaching (Danielson Group, 2013) specifically calls out designing coherent instruction as a necessary component of the lesson planning and preparation process. Coherent instruction means that the sequence of learning activities follows a coherent sequence, sensibly builds on one another, and is clearly in service of the instructional goals. Providing coherent instruction warrants a solid understanding of state, district, and school policy and expectations, as well as knowledge of content, standards,

and student needs. When the teacher uses formative and summative assessments that are linked to one another, the task of designing coherent lesson plans becomes less daunting because the interconnected assessments provide the teacher with congruous data that measure the same skill and knowledge set upon which to base adjustments in instruction. Black and Wiliam (2003) noted the need to “align formative and summative work in new overall systems, so teachers’ formative work would not be undermined by summative pressures, and indeed, so summative requirements might be better served by taking full advantage of improvements in teachers’ assessment work” (pp. 623-624).

Furthermore, a statistically significant relationship between the students’ reading proficiency (as measured by STAR Reading assessment) and their overall English Language Arts and Literacy academic achievement (as measured by the SBA English Language Arts/Literacy) may serve to illustrate the extent to which reading influences other important academic and language skills such as writing effectively, speaking clearly and coherently, listening actively, and engaging in research and inquiry. The results of this study may confirm for educators the important role that reading plays in so many areas of academics, and will hopefully spur them to strive to or continue to integrate reading into their instruction as much as possible.

From an organizational and leadership standpoint, this study has practical significance in that it may help educational leaders make decisions about professional development planning and allocation and redirection of funds in order to support teacher efforts to provide meaningful literacy instruction and targeted intervention and enrichment. This is especially important for minority and low income students who consistently lag behind their peers in reading and can certainly benefit from a responsive,

well thought out system of interventions that extend or supplement classroom-based intervention efforts.

In addition, school districts invest time, money, and human resources on administering and tracking results of a reading progress monitoring assessment such as STAR Reading. Thus, it would be helpful to examine the usefulness of the STAR Reading assessment results to understand if it is worth district and school investment of resources or if an alternative interim assessment that is better aligned with the state standards should be explored.

This study is also practically significant because of its focus on third grade, widely considered a pivotal year for literacy and reading skills. Research (Hernandez, 2011; Lesnick, Goerge, Smithgall, & Gwynne, 2010) has suggested that students reading proficiently in third grade matters greatly because it prepares them for complex reading tasks in fourth grade and beyond. It is, therefore, particularly important for teachers to improve their reading and literacy instruction in third grade to adequately prepare their students for more complex reading activities in the ensuing grade levels. Examining reading data of third grade students is a worthy exercise that provides schools with helpful information upon which to base instructional decisions and adjustments designed to provide differentiation, enrichment, and/or intervention. Moreover, careful examination of reading assessment data helps teachers address reading challenges before it is too late. Clay (1985) asserted that the reading challenges faced by a young child might be overcome more readily if the student had practiced error behavior less often and did not have very many reading habits and strategies to unlearn and/or relearn.

In response to federal and state mandates holding them accountable for students' reading and literacy skills, many school districts are using data from progress monitoring measures or benchmark assessments to inform teacher practice, identify students at risk of reading difficulties and address their reading challenges, and promote reading proficiency on state standards and state assessments. It is important to pay very close attention to students' literacy development because, according to the National Association for the Education of Young Children (1998), one of the best predictors of whether a child will function competently in school and go on to contribute actively in an increasingly literate society is "the level to which the child progresses in reading and writing" (p. 3). Therefore, a number of studies have been conducted to determine the efficacy of interim measures and classroom-based measurements and whether they predict students' mastery of literacy standards as measured by state exams (Miller, Bell, & McCallum, 2015; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; McGlinchey & Hixson, 2004; Weinstein, 2011; Wood, 2006). This study substantively contributes to the existing body of research by specifically focusing on the Smarter Balanced Assessment (SBA). The SBA is a new assessment used in the state of Washington and many other states across the country to measure students' mastery of the reading skills outlined in the Common Core State Standards. Few previous research studies on progress monitoring measures and state standardized exams use the Smarter Balanced Assessments as a variable. This study is also a significant addition to existing research because it involves STAR Reading, which not only tests students' foundational reading skills such as oral reading fluency, but also tests comprehension. Previous research on this topic primarily involved oral reading fluency measures such as DIBELS. The use of STAR Reading as

an independent variable in this study is significant because of a growing belief amongst educators that teaching reading comprehension does not need to be postponed until students are able to read fluently, a notion propelled by research studies suggesting young students benefit from instruction around reading comprehension (Reutzel, 2015).

The theoretical significance of this study includes the potential to validate the theory of Structural Cognitive Modifiable Theory (SCMT), which is centered on the belief that students' cognitive structures can be changed and that students' ability is not fixed (Feuerstein, 1990). This change in cognitive structure is possible through such approaches as mediated learning experience. Mediated learning experience is the way in which stimuli experienced in the environment are transformed by a mediating agent, such as a teacher, a coach, or a mentor, in the life of the learner (Feuerstein, Klein, & Tannenbaum, 1999). In the classroom, a teacher is a mediating agent whose principal job is to transform stimuli experienced by students in order to help them learn. Teachers can transform stimuli for the purposes of student learning through a variety of effective practices, such as those laid out by Danielson, Axtell, Bevan, Cleland, McKay, Phillips, and Wright (2009) in the Danielson Framework for Teaching. Such practices include the use of effective questioning and discussion techniques, collaborative work, appropriate lesson structure and pacing, and demonstrating flexibility and responsiveness. In addition, the effective use of assessment is critical to offering a meaningful mediated learning experience if student learning data can be analyzed and acted upon to provide learning experiences truly targeting students' needs, gaps, and deficiencies. Specific to reading and literacy, teachers can make better decisions as to what reading strategies to teach students (e.g., annotation and scanning for thesis statements, topic sentences, and

important ideas) and what reading skills to focus on (e.g., inferring, summarizing, vocabulary contextual clues, and reading fluency) for the purposes of improving students' literacy if they use data that indicate the students' reading gaps.

It is important to note that while classroom-based formative assessments are very useful in generating data about students' reading skills germane to a specific lesson, progress monitoring measures such as STAR Reading that assess skills and concepts addressed in the Common Core State Standards provide teachers with an even clearer picture of students' progress towards meeting the standard. They generate a broad range of data, both detailed and thorough, to guide crucial instructional decisions in the classroom (Renaissance Learning, 2011). Because they are short and efficient, they can be easily and readily incorporated into the instructional schedule without consuming too much instructional time. This allows the mediating agent, the teacher in this instance, to better mediate students' learning experiences so they can become proficient readers and academically successful.

Research Questions

The study considers the following research questions:

Question 1. Is there a statistically significant relationship between STAR Reading spring scores and SBA English Language Arts/Literacy assessment scores in third grade?

Hypothesis 1. There is a statistically significant relationship between STAR Reading spring scores and SBA English Language Arts/Literacy scores in third grade.

Question 2. Is there a statistically significant relationship between STAR Reading assessment scores from fall, winter, and spring and SBA English Language Arts/Literacy scores in third grade?

Hypothesis 2. There is a statistically significant relationship between the students' scores on the STAR Reading assessment from fall, winter, and spring and their SBA English Language Arts/Literacy scores.

Question 3. Is there a statistically significant relationship between students' performance on the SBA English Language Arts/Literacy and their STAR Reading spring score, gender, Free and Reduced Lunch status, and Special Education status in third grade?

Hypothesis 3. There is a statistically significant relationship between students' SBA English Language Arts/Literacy scores and their STAR Reading spring score, gender, Free and Reduced Lunch status, and Special Education status in third grade.

Research Methods

This study collected and analyzed preexisting 2014-2015 STAR Reading and SBA English Language Arts/Literacy data from 651 third grade students in a semirural, medium-sized school district in the Pacific Northwest. STAR Reading was administered three times during the 2014-2015 school year, during the fall, winter, and spring. The summative SBA English Language Arts/Literacy testing was administered in the spring of the 2014-2015 school year.

The statistical procedures used in the study were Spearman's rank-order correlation, hierarchical multiple regression, and standard multiple regression. To answer the first research question, the Spearman's rank-order correlation was used. Spearman's correlation is used to determine the strength and direction of the association or relationship between two continuous and/or ordinal variables (Field, 2013). For this

particular study, the STAR Reading spring score is an ordinal variable and the SBA English Language Arts/Literacy score is a continuous variable.

To answer the second research question, hierarchical multiple regression was conducted. Hierarchical multiple regression was used to assess the relationship between the fall, winter, and spring STAR Reading scores (independent variables) and the SBA English Language Arts/Literacy score (dependent variable). Furthermore, hierarchical multiple regression was conducted to determine if the STAR Reading fall and winter scores accounted for the variance in the SBA score over and above the STAR Reading spring scores.

To answer the third research question, standard multiple regression was utilized. Multiple regression was used to determine a statistically significant relationship between the dependent variable (SBA English Language Arts/Literacy) and multiple independent variables; namely, the STAR Reading spring scores, gender, socioeconomic status, and SPED status.

Terms and Definitions

Reading. According to Mooney, 1990, reading, is the creation and recreation of meaning, and it takes place through nonverbal as well as verbal modes of language—through listening and speaking, reading and writing, moving and watching, shaping and viewing. Reading is not merely a curriculum subject able to be confined to any one period, for reading is a part of any exchange of meaning through text. (pp. 2-3).

Early literacy. Early literacy refers to the knowledge, skills, and dispositions that students in primary grades, kindergarten through third grade, must possess in order to

learn to read and write. It is a dynamic process of forming reading and writing concepts and skills (Roskos, Christie, & Richgels, n.d.).

Literacy. Literacy is the process of learning about the print form of language and being able to use it in order to communicate (Connecticut State Department, n. d.).

Reading fluency. Reading fluency is a complex reading construct that involves not only speed and accuracy but also prosody or inflection (Miller, Bell, & McCallum, 2015).

Summative assessment. Summative assessments typically are administered at the end of a unit of time such as the end of the school year in order to gauge student mastery of content standards. “These assessments typically are given statewide (but can be national or district) and these days are usually used a part of an accountability program or otherwise inform policy” (Perie, Marion, Gong, & Wurtzel, 2007, p. 1).

Formative assessment. Formative assessments are used by teachers and students during instruction that provide feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes...The assessment is embedded within the learning activity and linked directly to the current unit of instruction” (Perie et al., 2007, p. 7).

Interim assessment. Interim assessments fall between formative and summative assessment and “may be given at the classroom level to provide information for the teacher, but unlike true formative assessments, the results of interim assessments can be meaningfully aggregated and reported at a broader level” (Perie et al., 2007, p. 1).

Progress monitoring measures. Progress monitoring measures are interim assessments that are short, efficient, frequent assessments to track growth rate as well as

level. It increases in frequency as the need for intervention increases, although the ideal system provides for continuous progress monitoring so that robust series of data are always available (Renaissance Learning, 2011). This term is sometimes interchangeable with benchmark assessment or curriculum-based measurement.

Benchmark assessment. A benchmark assessment is an interim assessment that can be used either formatively or summatively (Henderson, Petrosino, Guckenberger, & Hamilton, 2007). Henderson et al. (2007) state that it “provides local accountability data on identified learning standards for district review after a defined instructional period and provides teachers with student outcome data to inform instructional practice and intervention before annual state summative assessments” and enables “educators to monitor the progress of students against the state standards and to predict performance on state exams” (p. 2).

Curriculum-based measurement. Curriculum-based measurement is comprised of formative assessments that allow teachers to make informed instructional decisions through regular and continued monitoring of student growth in core academic areas such as reading, writing, and math (Deno, 1985; Howell & Nolet, 1999).

Chapter Two

Literature Review

The value of learning to read cannot be overstated. Learning to read makes reading to learn possible, which then enables the reader to enrich his/her perspectives, reflect on differing viewpoints, stay current on what is happening around the world, broaden his/her knowledge base, and question existing beliefs and values, amongst many other noteworthy benefits. Literacy is unequivocally a requisite skill in today's competitive global economy. Thus, schools across the country make literacy and reading, especially in light of federal and state mandates and legislation, a top priority and goal around which continuous improvement plans, strategic direction, professional development efforts, and accountability measures revolve. Elementary schools, in particular, must work hard to promote early literacy so students reach acceptable levels of reading proficiency by the time they leave third grade, lest students are placed at risk of long-term academic and life struggles. In efforts to promote reading and literacy, many schools administer benchmark measures or progress monitoring measures to inform reading instruction. The following research review consists of two parts: (a) theoretical frameworks and (b) review of empirical research. The theoretical frameworks discuss a learning theory as it pertains to the use of progress monitoring measures to help students achieve reading proficiency. The review of empirical research discusses relevant empirical studies that (a) are linked to reading proficiency in third grade, (b) illustrate the impact of reading proficiency on other aspects of learning, and (c) highlight the use of progress monitoring assessment in predicting student performance on standardized state tests.

Theoretical Framework

Using progress monitoring assessments such as STAR Reading in tracking and monitoring students' reading skill level and growth is aligned with the theory called Structural Cognitive Modifiability.

Structural Cognitive Modifiability Theory. Structural Cognitive Modifiability Theory (SCMT), developed by Feuerstein (1990), rejects the argument that certain human conditions, such as cognitive capacity, are irreversible. Insisting human beings are not unmodifiable, the theory, according to Feuerstein (2008), counters the notion that the mental conditions of the individual were... irreversible, unchangeable—human beings born with a given level of functioning will have to finish their lives with the level they were born with, irrespective of what they may have achieved over the years”—a school of thought he considers to be a “very inappropriate influence on education. (p. 5)

The theory of Structural Cognitive Modifiability instead advances the idea that cognitive capabilities are a dynamic, flexible construct and that people have the potential to change cognitively at all stages of development (Feuerstein, Klein, Tannenbaum, 1999). According to Feuerstein et al. (1999), the term “structural” refers to the organization and integration of the different components that comprise the way we think, while the term “cognitive” means the ability to learn, reason, and think. Modifiability, on the other hand, describes one's ability to adapt and regulate (Feuerstein et al., 1999). All three terms taken together mean that all learners have the potential to change, adapt, or regulate the way that they think, learn, construct meaning, and apply various skills in a specific context. The Structural Cognitive Modifiability Theory underpins the practice of

using formative and interim assessments in instruction, which does not focus on the end product but on the process of learning and effecting growth. The use of formative and interim assessment data is an acknowledgement that students' cognitive ability is not fixed, and that given appropriate, data-informed instruction, intervention, reflection, or learning interactions, a student can demonstrate academic growth.

Structural Cognitive Modifiability (Feuerstein et al., 1999) is made possible by such approaches as “mediated learning experience,” which refers to the way in which a mediating agent (i.e., a parent, teacher, mentor, or coach) transforms the environmental stimuli experienced by the learner. The mediating agent, steered by intention, student learning goals, emotional investment, and culture, influences, enhances, or organizes the world of stimuli in a learner's life in order to make it conducive to learning according to articulated learning goals, (Feuerstein et al., 1999). Feuerstein, Falik, and Feuerstein (2003) explained that for mediated learning experience to happen,

an *intentional* human being must interpose him or herself between the stimuli and the learner's response to the learning. This is mediation in the sense that the situation (stimuli and responses) are modified by affecting qualities of intensity, context, frequency, and order, while at the same time arousing the individual's vigilance, awareness, and sensitivity. The interactional experience may have the quality of repeating or eliminating various stimuli, relating events in time or space, or imbuing experience with meaning. (p. 54)

Feuerstein et al. (2003) postulated that an intentional mediator must “make planned and systematic choices to exploit the mediational potential of the situation to encourage cognitive functioning and stimulate modifiability” (p. 54). Thus, the

mediational experience must create a closed loop between the mediator, the mediatee, and the message or content of the interaction (Feuerstein et al., 2003). The concept of an intentional mediating agent aligns with Vygotsky's (1978) Zone of Proximal Development, which is described as the "distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 33). The Zone of Proximal Development supports the theory offered by Feuerstein in that it underscores the important role that an experienced adult or peer, or a mediating agent in Feuerstein's words, plays in a student's learning process.

Mediated learning experience is also rooted in the active modification approach, which views the past as merely a starting point for improvements in the future. Applying this concept in the classroom, the active modification approach, therefore, involves having a very clear understanding, which can be achieved through the use and analysis of assessment data, of students' academic entry point and then leveraging that knowledge to bring about growth and improvement. Pellegrino (2003) declared "in educational assessment, the information collected is designed to help teachers, administrators, policy makers, and the public infer what students know and how well they know it, presumably for the purpose of enhancing future outcomes" (p. 48). According to the Structural Cognitive Modifiability Theory, it is indeed possible to bring about the necessary modification to the learners' cognitive structures in order to generate better academic results.

The Structural Cognitive Modifiability Theory is applicable and relevant to this present study because the use of a progress monitoring measure is very much aligned

with the concepts of mediated learning experience and the active modification approach. The teacher administers the progress monitoring assessment as frequently as needed, and in response to the results, the teacher mediates the learning experience using the active modification approach by providing data-informed and targeted learning opportunities. This series of events helps address specific academic deficiencies and learning gaps in order to get students from their academic starting point to an intended new place in their learning. Feuerstein et al. (2003) asserted that mediated learning experiences are “animated by intentionality” (p. 54). This intentionality requires the mediator to be “alert, vigilant, and animated if the situation is to have all the necessary conditions to assure that the subject grasps the task and is ready to focus and interact with it (Feuerstein et al., 2003, p. 54). It is nearly impossible to be “animated by intentionality” and provide intentional, vigilant, animated, well-planned, and systematic mediated learning experiences absent any assessment data.

Progress monitoring measures are a practical, feasible, and useful tool that teachers can use to mediate students’ learning experiences in the classroom because these types of assessments are short and efficient (Renaissance Learning, 2011). It is manageable for teachers to tightly and regularly incorporate progress monitoring measures into the curriculum because they take minimal time to administer and yet offer detailed and thorough data about student progress toward learning goals and academic standards. In the area of reading and literacy, teachers can use progress monitoring measures and take appropriate, intentional, and targeted action based on the test results in order to change a struggling student’s reading proficiency trajectory.

In sum, if students enter third grade not reading proficiently, educators must do all they can to mediate these students' learning experiences and address their reading gaps, understanding it is possible for students to improve their skills with the appropriate support and intervention. Progress monitoring measures are powerful, efficient, and practical tools that can be utilized towards that end. Neglecting to offer effective, data-informed intervention to struggling readers may lead to significant academic challenges and consequences as discussed in the following empirical research on third grade reading.

Empirical Research on Third Grade Reading

There is a sense of urgency for educators to ensure young students become proficient readers by the time they finish third grade, given that fourth grade begins to expose students to even more complex texts and tasks that require a higher level of comprehension. Below are research studies that discuss why reading proficiency by the end of third grade is critical.

Not being able to read proficiently by the end of third grade carries with it dire educational consequences such as failure to graduate from high school on time. Hernandez (2011) conducted a longitudinal study that calculated the graduation rates of students with various reading abilities and varying poverty rates. The study examined database records of 3,975 students who were born from 1979 to 1989 in efforts to decipher the impact of third grade reading skills and poverty on high school graduation.

The researcher categorized the students into three income groupings: (a) those who have never lived in poverty, (b) those who spent some time in poverty, and (c) those who have lived more than half the years surveyed (five years) in poverty. Every two years, the parents of the participating students were surveyed in order to determine the

family's socioeconomic status. The students were also separated into three reading groups, according to the following reading levels: proficient, basic, and below basic (Hernandez, 2011). The reading progress of the participating students, according to the author, was monitored through the yearly administration of the Peabody Individual Achievement Test (PIAT).

The database records showed whether the students had finished high school by age 19. Analysis of these records revealed that 88% of the students graduated from high school by age 19. However, graduation rates differed dramatically for students with varying reading skills in third grade. According to Hernandez (2011), one in every six students who were not reading proficiently in third grade did not graduate from high school on time. Only 4% of students who were proficient readers in third grade failed to graduate on time, while 16% of students who were reading below grade level missed the on-time graduation mark. Among students who were not proficient readers in third grade and did not graduate from high school on time, 9% had basic reading skills and 23% were considered to have below basic reading skills (Hernandez, 2011).

Students who were not proficient readers in third grade and also lived in poverty were approximately three times more likely to drop out or fail to graduate from high school by age 19 compared to their peers who have never lived in poverty (Hernandez, 2011). The author added that students from low-income families typically do not have access to decent housing, food, clothing and books; neither do they generally have access to high quality childcare, health care, or early education. These lack of resources unfortunately translated into the students entering kindergarten with weak academic skills and without the foundational language or social skills needed for academic success.

Students who do not read proficiently by the end of third grade also generally do not catch up with their peers and continue to struggle with reading difficulties in later years. This challenge underscores the importance of providing timely, effective, and data-informed intervention practices before it is too late. A longitudinal analysis of the educational outcomes of third grade students enrolled in Chicago schools in the 1996-1997 school year found third grade reading level to be a significant predictor of eighth grade reading level (Lesnick et al., 2010). Third grade reading level did not entirely influence eighth grade academic outcomes; however, the authors reported a strong correlation between third grade and eighth grade reading levels ($r = 0.67$; no p -values and other pertinent correlational data were reported). Using multilevel regression models that take into account third grade school effects and demographic characteristics of students, the researchers found that students who were at or above grade level in reading as third graders were more likely to read at or above grade level as eighth graders. Among those who were reading below grade level in third grade, approximately 40% were also reading below grade level in eighth grade. For students who were reading at grade level as third graders, their third grade reading level did not have an impact on their eighth grade reading level. The research also showed that those who were reading above grade level were more likely to attend college than their peers, even after taking into account demographics, eighth grade reading level, ninth grade school effects, and course performance. This study's findings reinforced the need for struggling readers to receive timely, targeted, responsive, and intentional support and intervention in order to help them make progress towards high school graduation and college admission.

Another consequence that results from reading struggles during primary years is related to social skills and academic achievement. Miles and Stipek (2006) found a connection between poor literacy achievement in the first grade and subsequent development of aggressive behavior. This confirmed the necessity of effective literacy instruction in a student's early years of school. To investigate the relationship between literacy achievement and social skills, specifically aggression and prosocial behavior, the researchers conducted a longitudinal study involving low-income children ($N = 400$) at three K-5 elementary schools in the Northeast and on the West Coast. Data were collected on two cohorts of students, aged four to six years old when the study began; one cohort was entering kindergarten and the other group entering first grade. Both groups were assessed again in third grade and fifth grade respectively. To measure social skills, teachers were asked to rate the students' aggressive and positive social behavior. Literacy achievement was measured using reading, comprehension, and writing tests.

The researchers calculated bivariate correlations to measure associations between the participants' literacy achievement and their social behavior, such as aggression and prosocial behavior. Results revealed that literacy in the first grade had a statistically significant, negative correlation to aggression in fifth grade ($r = -.28, p < .01$) and third grade ($r = -.32, p < .01$). First grade prosocial behavior was also statistically significantly associated with literacy in fifth grade ($r = .24, p < .05$) and in third grade ($r = .24, p < .05$).

Hierarchical linear regression analysis, according to Miles and Stipek (2006), was used to

test the hypothesis that (a) the effect of first grade literacy achievement on fifth grade aggression was mediated by the effect of first grade literacy achievement on third grade aggression and (b) that the effect of first grade prosocial behavior on fifth grade literacy achievement was mediated by the effect of first grade prosocial behavior on third grade literacy achievement. (p. 111)

According to Miles and Stipek (2006), first grade literacy achievement did not significantly predict fifth grade aggression when third grade aggression was added to the regression model ($\beta = -.07, p = .55$). In the same vein, first grade prosocial behavior did not significantly predict fifth grade literacy achievement when third grade literacy achievement was added to the regression model ($\beta = -.09, p = .34$). The authors believed their findings, which demonstrated the link between social skills and academics, highlighted the significance of promoting both academic and social skills in the elementary school years.

Empirical Research on the Impact of Reading on Academic Success

Literacy experts and scholars claim reading is strongly associated with student learning and academic success (Lonigan & Phillips, 2015; Nation & Norbury, 2005; O'Connor & Klein, 2004). Reading requires the ability to synthesize, construct meaning, comprehend, analyze, and evaluate, which are higher-order thinking skills that are transferrable to any learning experience in the classroom and critical to academic success. The following research studies were conducted to explore the relationship between reading proficiency and academic achievement. They affirmed the commonly held belief that a student's reading ability does have an impact on their academic achievement, be it in the area of math, science, or English Language Arts. These research studies are

relevant because they offer the present study with empirical backing to support the hypothesis that students' reading proficiency, as measured by STAR Reading, is linked to students' overall literacy (reading, writing, speaking and listening, research and inquiry) achievement, as measured by the SBA.

Espin and Deno (1993) conducted a study involving 121 10th grade students in a rural school in the Midwest. The study investigated the relationship between reading and literacy skills and academic achievement. The researchers found that there was a moderate to high relationship between reading abilities and student academic success. For instance, standardized reading achievement results were moderately correlated to grade point average ($r = .56, p < .001$) while reading achievement results were highly correlated with overall math proficiency ($r = .70, p < .001$), science proficiency ($r = .74, p < .001$), social studies achievement ($r = .77, p < .001$), and written expression ($r = .60, p < .001$).

Another study claimed that higher reading comprehension leads to higher achievement in science (Cromley, 2009). Examining the relationship between scientific literacy and reading literacy, the researcher used three international data sets from the Programme on International Student Assessment (PISA), which was administered to 15 year-old students in 2000, 2003, and 2006. Correlation between scientific literacy and reading literacy was calculated. Cromley (2009) found a strong correlation between scientific literacy and reading literacy across all data sets ($r = .840, p < .001$ for the 2000 PISA results; $r = .805, p < .001$ for the 2003 PISA results; $r = .819, p < .001$ for the 2006 PISA results).

Research has also found an association between reading proficiency and math achievement. One study conducted by Vilenius-Tuohimaa, Aunola, and Nurmi (2008) explored the relationship between reading comprehension and mathematical word problem skills of 225 fourth graders in Central Finland. The researchers found that students' ability to solve math word problems was correlated to reading comprehension. Results of the Pearson correlation showed that reading comprehension variables were interrelated with math word problem variables. For example, the reading comprehension skill of conclusion and interpretation and the math word problem skill around change were strongly correlated ($r = .05, p < .001$). The reading comprehension skill of identifying cause and effect and structure was moderately but statistically significantly correlated with the math word problem skill of combining ($r = .46, p < .001$). The authors concluded that both reading and math word problem solving required reasoning abilities.

Experts also claim reading and writing are interdependent and that reading positively impacts a student's writing skills. Fletcher and Portalupi (1998) believed the writing classroom is built on the foundation of literature, asserting that the written work that teachers get out of their students "...can only be as good as the classroom literature that surrounds and sustains it" (p. 10). Loban (1963) stated "those who read well also write well; those who read poorly also write poorly" (p. 75). Olness (2005) maintained one reason to expose children to quality literature is its influence on student writing, with students, either consciously or unconsciously, using literary models when they write. He explained that when reading a story,

students hear the language of good writers, are exposed to rich vocabulary, and develop literary awareness, or 'a sense of story.' They learn the structure and

language of books. And they acquire literacy skills that can be transferred to their own writing. (p. 2)

Stotsky (1983) reviewed correlational and experimental studies from the 1930s to 1981 that investigated the relationship between students' reading and writing skills. Findings showed "better readers tend to produce more syntactically mature writing than poorer readers" (p. 636). The author also cited studies that demonstrated how reading experiences may be as good as, if not better than, grammar study and additional writing practice in enhancing students' writing skills.

The present study aims to investigate the relationship between third grade students' reading proficiency level as measured by STAR Reading and their overall literacy achievement as measured by the Smarter Balanced Assessment. The above mentioned studies are relevant to this purpose because they provide research-based evidence that affirms reading proficiency has a relationship to students' academic achievement.

Empirical Research on Benchmark Measurements

To help address deficiencies and mediate learning experiences in third grade so students can become proficient readers, interim assessments are necessary. They provide teachers with actionable data to help promote literacy among students, which, in the case of Washington State, is summatively assessed using SBA English Language Arts/Literacy. The following research studies were conducted to examine the relationship between interim measurements and standardized tests. They are relevant to the current study because their findings offer insights into the value and utility of interim

assessments in predicting student performance on standardized exams, which is what this current study aims to do.

A study conducted by Miller et al. (2015) compared student scores on a CBM or classroom-based measurement (Monitoring Instructional Responsiveness: Reading or MIR:R) to their scores on the Tennessee Comprehensive Assessment Program Reading Composite or TCAP Reading.

Third grade students ($N = 448$) enrolled in a rural school in the southeastern United States participated in the study. School demographics were not identified; however, information about the school district was furnished. Nearly 60% of the student population across the district were from low income families, and 95% of the students were Caucasian.

The MIR:R, administered to a group of students in three minutes, was comprised of four passages, with 10 sentences each, which were both expository and literary (Miller et al., 2015). According to the authors, the passages did not have any type of punctuation marks nor capital letters, and students were required to determine where one idea ends and another starts when reading, signifying it with a slash mark. The MIR:R measured the students' comprehension percentage, total words read, and comprehension rate. Comprehension rate is considered to be a function of total words read (rate) and comprehension percentage (Miller et al., 2015).

The TCAP, on the other hand, is a timed, criterion referenced standardized exam administered in the spring of each school year to third and eighth graders in the state of Tennessee (Miller et al., 2015). The exam assesses students in reading, language, arts, mathematics, science, and social studies. TCAP reports raw scores and scaled scores for

each of the content areas tested. This specific study used the scores from the TCAP reading composite, which included 39 questions.

Calculating Pearson product-moment correlation, the researchers found that the relationship between the MIR:R Comprehension Rate score and the TCAP reading score was moderately strong ($r = .58; p < .01$). Comprehension Percentage and TCAP also had a moderately strong relationship to one another ($r = .54; p < .01$). Total Words Read, on the other hand, did not have a statistically significant relationship to TCAP ($r = -.01; p > .05$).

The researchers used stepwise multiple regression to ascertain the relationship between two MIR:R component scores (Comprehension Percentage and Total Words Read) and TCAP, with Comprehension Percentage placed in the equation first. Miller et al. (2015) reported that both component scores showed predictive utility. The MIR:R Comprehension Percentage score predicted 29% of the variance in TCAP scores ($R^2 = .29; p < .001$). The MIR:R Total Words Read predicted a meager additional 1% of the variance in TCAP scores ($R^2 = .01; p < .05$). When Total Words Read was combined with Comprehension Rate, the combined scores predicted 35% of the variance in TCAP scores ($R^2 = .35; p < .001$).

Miller et al. (2015) claimed that the study's findings offered proof that an efficient, multifaceted CBM tool for reading, such as MIR:R, "can yield a score that is related to a high stakes, end of year test" (p. 715). This allows teachers to use CBMs like the MIR:R with some assurance that they can help identify at-risk students in need of intervention, especially if the data is obtained early enough in the year to give teachers time to be responsive to the data (Miller et al., 2015).

Wood (2006) investigated the relationship between oral reading fluency, using DIBELS, and student performance on the reading component of the Colorado Student Assessment Program (CSAP). A total of 281 students in third grade ($n = 82$), fourth grade ($n = 101$), and fifth grade ($n = 98$) enrolled in a public school in northern Colorado participated in the study. Approximately 11% of the participants received Special Education Services, with 81% of them receiving support in reading. A majority of the participating students were Caucasians (89% in third grade, 85% in fourth grade, and 84% in fifth grade), while Hispanic students comprised 10% of the participants at each grade level. Native Americans represented roughly 1% of the participating students.

Each year, third grade students in Colorado take the CSAP in the month of February, while fourth and fifth graders take it in March (Wood, 2006). The researcher explained that the reading component of the CSAP assesses students' mastery of the state standards, specifically student understanding of a variety of materials, application of thinking skills to reading, making use of relevant information, and recognizing literature as a record of human experience. Wood (2006) added that these same content standards are measured each year even though the questions, multiple choice and constructed response, vary from year to year. Student performance is grouped into four levels, namely, advanced, proficient, partially proficient, and unsatisfactory, and is reported in scale scores (Wood, 2006).

According to Wood (2006), the DIBELS Oral Reading Fluency measure contains three benchmark passages for each grade level and is administered in the fall, winter, and spring. The researcher stated that for this particular study, students were asked to read from each of the three passages for one minute, and the number of words read correctly is

recorded as their oral reading fluency scores. Wood (2006) reported that students were tested for one week on the DIBELS Oral Reading Fluency two months in advance of the CSAP and that the median score was recorded and used for comparison with the CSAP scores. The Pearson's r , according to Wood (2006), showed significant correlations between DIBELS and CSAP for all three grade levels ($r = .70, p < .001$ for third grade; $r = .67, p < .001$ for fourth grade; and $r = .75, p < .001$ for fifth grade).

To determine whether or not DIBELS Oral Reading Fluency added to the predictability of performance on the CSAP over and above previous years' performance, Wood (2006) employed multiple regression with fourth grade and fifth grade data. The fourth grade CSAP score was the dependent variable, while the predictor variables were the third grade CSAP score and fourth grade DIBELS Oral Reading Fluency score. The same analysis was used for fifth grade with the CSAP fifth grade score as the dependent variable and the fourth grade CSAP score and fifth grade DIBELS score as the independent variables. Wood (2006) wrote that regression analysis showed that both the fourth grade DIBELS score and the third grade CSAP score were significant and independent predictors of fourth grade CSAP scores, accounting for 62% of the variance in the dependent variable ($R^2 = .62, p < .001$). Similar to the fourth grade results, the fifth grade DIBELS score and fourth grade CSAP score were significant and independent predictors of fifth grade CSAP scores ($R^2 = .70, p < .001$). Based on these findings, Wood (2006) found a statistically significant, strong relationship between oral reading fluency and student performance on statewide reading proficiency tests. Wood (2006) posited that the study's findings are significant in that they provide schools districts

with additional information about individual students from oral reading fluency measures even after prior performance on a statewide test is considered. This has the potential to further improve identification of needs, instructional planning, and intervention of students at different reading levels. (p. 101)

Another study inquired into the relationship between students' performance on two progress monitoring measures (oral reading fluency or ORF: DIBELS or running records; reading comprehension: 4sight Assessment) and the Pennsylvania Systems of School Assessment (Weinstein, 2011). The researcher reported that two cohorts of students, third grade ($n = 205$) and fourth grade ($n = 171$) students enrolled in four suburban elementary schools in 2009-2010, participated in the study. A majority of the younger cohort of third graders were Caucasians (91.2%), while 8.8% were African Americans. Hispanics, multiracial, and American Indians made up the rest of the student population. Nearly 30% of the students came from low income families, and 19.5% qualified for Special Education services. Among the older cohort of fourth graders, 23.4% of students qualified for Special Education services, and 26.9% of students qualified for Free and Reduced Lunch. Similar to the third grade cohort, a majority of the fourth graders were Caucasian (96.5%), while 3.5% were minority students.

Weinstein (2011) indicated that to measure students' oral reading fluency, DIBELS data from fall, winter, and spring were analyzed. The DIBELS recorded the number of words students read correctly in one minute. The author added that in the absence of DIBELS data, running records scores, which also counted the number of words read per minute, were used. Both DIBELS and running records had the same three oral fluency categories: low risk, some risk, or at risk (Weinstein, 2011).

The Pennsylvania System of School Assessments (PSSA) is given to all students each spring; the reading portion of which measures students' reading achievement (Weinstein, 2011). The 4Sight Assessment, Weinstein (2011) noted, mimics the actual PSSA test and was designed for the purpose of predicting student scores on the Pennsylvania state test. Both the 4Sight Assessment and PSSA have the same level of categories (advanced, proficient, basic, and below basic). Students in the participating elementary schools took the reading 4Sights Assessment in September, November, and February (Weinstein, 2011).

Archival data, which included DIBELS or running records oral reading fluency scores, 4Sight Assessment scores, and PSSA scores, were analyzed for both groups of students participating in the study. Weinstein (2011) used correlations to establish the relationship between the PSSA and the oral reading fluency measures and the reading monitoring benchmark. The researcher reported that the 4Sights Assessments produced higher correlations with the PSSA than the oral reading fluency measures ($r = .61$ for ORF vs. $r = .77$ for 4Sight in 3rd grade in the spring time; $r = .48$ for ORF vs. $r = .67$ for 4Sight for fourth grade in the spring time; no p -values were reported). Weinstein (2011) concluded that reading comprehension benchmarks were, according to study findings, better indicators of PSSA than oral reading fluency measures. It must be noted that because statistical significance was not reported, this result needs to be interpreted with caution. Nevertheless, Weinstein (2011) believed school districts are better served to place stronger emphasis on reading comprehension benchmark measures than oral reading fluency measures when predicting student performance on a state standardized exam such as the PSSA.

Shapiro, Keller, Lutz, Santoro, and Hintze (2006) conducted a study to explore the relationship between curriculum-based measures and performance on standardized tests in reading, math computation, and math concepts/applications in two school districts in Pennsylvania. For the purposes of this current study, only the results pertaining to reading will be discussed.

Shapiro et al. (2006) reported that third, fourth, and fifth grade students ($N = 2,938$) in two school districts in eastern Pennsylvania were involved in the study. The two school districts were moderately sized with a combination of urban and suburban schools. The average percentage of students, according to the researchers, who came from low-income families was approximately 20%.

To examine the relationship between progress monitoring and standardized assessment, Shapiro et al. (2006) used two types of assessments: (a) curriculum-based measures or CBM for reading and (b) standardized state assessment (PSSA). The CBM recorded the number of words that students read per minute from grade-based narrative reading passages (Shapiro et al., 2006). Archival CBM data collected from students in October, February, and May as part of the participating district's norming projects were analyzed for this study.

On the other hand, the PSSA reading test, according to the researchers, measures the following reading skills: "(a) learning to read independently; (b) reading critically; (c) reading, analyzing, and interpreting literature; (d) characteristics and function of the English language; and (e) research" (Shapiro et al., 2006, p. 24). Student scores in the PSSA, according to Shapiro et al. (2006), are grouped according to the following levels: below basic, basic, proficient, and advanced.

The researchers used Pearson product-moment correlation to examine the relationship between the PSSA and the CBM reading (oral reading fluency or ORF) scores obtained at fall, winter, and spring assessments in both districts in third, fourth, and fifth grades. Shapiro et al. (2006) reported statistically significant, strong correlations between the CBM fall, winter, and spring scores and the PSSA scores in both districts, with correlation coefficients ranging from .62 to .69, $p < .001$. The only correlation lower than .62 is between CBM fall and PSSA in District 2 ($r = .24$, $p < .001$). An interesting finding to note is that the hierarchical regression analysis showed that the winter assessment period was the strongest predictor of the students' PSSA performance, and that the spring assessments did not add significantly to the explanation of variance that contributes to the PSSA student scores (Shapiro et al., 2006). The researchers concluded that these findings have significance in that the CBM may be a helpful source of information with possible utility in identifying students who are at risk of not passing the statewide assessment and providing them with targeted intervention to address whatever learning gaps they may display.

McGlinchey and Hixson (2004) also examined the usefulness of curriculum-based measurement or CBM in predicting student performance on the Michigan Educational Assessment Program's (MEAP) fourth grade reading assessment in an eight-year longitudinal study. An elementary school in an urban school district in Michigan participated in the study for seven out of the eight years (1994-2001), while all fourth graders in the entire district participated in the study during year four (1997-1998). The researchers offered no explanation for the year four district-wide participation.

The elementary school that participated in the study served 450 to 520 students from kindergarten through sixth grade (McGlinchey & Hixson, 2004). The participating students in the elementary school included both general education and special education students, ranging from 55 to 139 in number each year. The researchers added that the school district had a student population of 11,000 students. Approximately 60% of the students in the district qualified for free and reduced lunch, and 52% of the student population were non-Caucasian. According to McGlinchey and Hixson (2004), throughout the eight-year study, participants numbered 1,362.

The study aimed to analyze the predictive value of an oral reading fluency CBM as it pertains to student performance on Michigan's state reading assessment, the MEAP (McGlinchey & Hixson, 2004). The researchers reported that the CBM used in the study was the Macmillan Connections Reading Program, a basal fourth grade reading test used by the district. Using passages randomly selected from the Macmillan Connections Reading Program, students read one passage out loud for one minute in the first five years of the study. This was due to time constraints and the large number of participants. With the increase of staff support during the last three years of the study, three one-minute reading probes, instead of just one, were used. The number of correct words per minute were recorded. McGlinchey and Hixson (2004) stated that a team of school staff made up of school psychologists, paraprofessionals, and school psychology interns were trained to administer and score the CBM reading tests over the eight-year period of the study.

The MEAP, a state approved assessment, assesses reading, writing, math, science, and social studies in fourth, seventh, and eleventh grades (McGlinchey & Hixson, 2004).

For this study, the researchers used the reading portion of the MEAP, an untimed test administered in a group setting over a period of two days. The test measures students' comprehension of literary and informational texts.

The study's findings indicated a moderately strong relationship between oral reading rates and MEAP performance. Correlation coefficients were consistent across all eight years of the study, ranging from .63 to .81 ($p < .001$), except for 1998-1999 ($r = .49, p < .001$). Using diagnostic efficiency statistics, the researchers reported that 72% of students who read at least 100 words correct per minute passed the state test. McGlinchey and Hixson (2004) concluded that a "simple, efficient, and repeatable assessment" (p. 202) such as an oral reading fluency CBM can predict student performance on a state test, which teachers, interested in ways to measure student reading progress to inform their practice, should find helpful. The researchers also maintained that because of the relationship between the CBM and state assessment, teachers can feel confident in using CBMs to help their students prepare for state assessments, which may compel school districts to adopt an empirically supported and efficient assessment practice.

The primary purpose of this present study is to examine the relationship between STAR Reading as a progress monitoring measure and the SBA English Language Arts/Literacy, which is a state standardized summative exam. The above mentioned studies are relevant to this purpose because they provide research-based evidence that depict a relationship between a progress monitoring measure or a benchmark assessment and a standardized summative test.

Chapter Three

Methods

The purpose of this study is to examine the relationship between student performance on a progress monitoring measure (STAR Reading) and the Washington State standardized exam (SBA English Language Arts/Literacy). This chapter contains the methods for the study, including the research questions, research design, participants, procedures, instrumentation, and data analysis.

Research Design

Multiple quantitative methods that include Spearman's rank-order correlation, hierarchical multiple regression, and standard multiple regression were used to answer the following research questions.

Question 1. Is there a statistically significant relationship between STAR Reading spring scores and SBA English Language Arts/Literacy assessment scores in third grade?

Hypothesis 1. There is a statistically significant relationship between STAR Reading spring scores and SBA English Language Arts/Literacy scores in third grade.

Question 2. Is there a statistically significant relationship between STAR Reading assessment scores from fall, winter, and spring and the SBA English Language Arts/Literacy scores in third grade?

Hypothesis 2. There is a statistically significant relationship between the students' scores on the STAR Reading assessment from fall, winter, and spring and their SBA English Language Arts/Literacy scores.

Question 3. Is there a statistically significant relationship between students' performance on the SBA English Language Arts/Literacy and their STAR Reading spring

score, gender, Free and Reduced Lunch status, and Special Education status in third grade?

Hypothesis 3. There is a statistically significant relationship between students' SBA English Language Arts/Literacy scores and their STAR Reading spring score, gender, Free and Reduced Lunch status, and Special Education status in third grade.

Participants

Convenience sampling was utilized in this research, with a medium-sized school district located in a semirural community in Washington State participating in the study. The school district has nearly 10,000 students. The school district's demographics in 2014-2015 do not fully mirror that of the state's as depicted in Table 1.

Table 1

Washington State and District Demographics

	Washington State	School District
Hispanic	21.7%	9.7%
American Indian/Alaskan Native	1.5%	0.7%
Asian	7.2%	3.8%
Black/African American	4.5%	1.1%
White	57.0%	78.9%
Two or More Races	7.1%	5.5%
Free and Reduced Lunch	45.0%	21.3%
Special Education	13.4%	13.9%
Transitional Bilingual	10.4%	3.1%

The schools in the participating school district include one primary K-2 school, nine K-6 elementary schools, one 3-6 elementary school, two middle schools serving seventh and eighth graders, two high schools, and one alternative high school. The district's on-time graduation rate is 85%, and its five-year graduation rate is 88%.

Preexisting data on the SBA and STAR Reading assessments from the district's 2014-2015 third grade cohort ($N = 651$) were analyzed in this study. The students in the 2014-2015 third grade cohort included roughly 50% male and 50% female. Over 75% of the students were Caucasian, 9% were Hispanic, and 9% were Two or More Races. Rounding out the rest of the student population were 4% Asian and less than 1% African American and Native American. Only 4% of students were considered English Language Learners, while 16% qualified for Special Education services. About 20% of the third graders were low-income students as determined by their qualification for Free and Reduced Lunch.

Instrumentation

STAR Reading. The STAR Reading assessment used in this study is an online assessment program for K-12 students that measures five reading domains: word knowledge and skills, comprehension strategies and constructing meaning, analyzing literary text, understanding author's craft, and analyzing argument and evaluating informational text (Renaissance Learning, 2015a). Appendix A shows the specific reading skills and domains that STAR Reading tests. It is a computer adaptive test, which means that the difficulty of the test is continually adjusted as the test progresses according to accuracy, or lack thereof, of the student's response to the previous question. The test contains three practice questions to ensure that the student knows how to

navigate the test and use the program. STAR Reading has over 5,000 questions in its item bank; 34 items are administered at each testing event, which takes approximately 15 minutes to complete (Renaissance Learning, 2014). It gives students one minute to answer each question before the program automatically moves on to the next question. STAR Reading uses a combination of traditional reading comprehension passages and questions (multiple choice) and cloze method. Cloze method is a type of reading comprehension question wherein every *n*th (e.g., seventh) word is replaced with a blank space, and a choice of three or more words are provided; the reader is then required to read the passage quietly and to fill in the blanks by selecting the correct word from available choices (Oller, 1979; Runge, Lillenstein, & Kovaleski, 2016). With the cloze procedure, the reader “not only reads the text but must also produce a word to fit a given context” (Raymond, 1988, p. 91).

In third grade, STAR Reading measures foundational literacy skills; namely, phonics and word recognition, fluency, comprehension, and language (Renaissance Learning, n.d.). It also measures comprehension skills, including key ideas and details, craft and structure, and integration of knowledge and ideas. Specific language skills tested are word relationships, vocabulary acquisition, structural analysis, and context clues.

The STAR Reading assessment in third grade aligns with the foundational reading skills and strategies identified by the National Early Literacy Panel (2008), which they call conventional literacy skills: decoding, oral reading fluency, reading comprehension, writing, and spelling. These skills, not including writing and spelling, are addressed in the STAR Reading test.

According to Renaissance Learning (n.d.), STAR Reading in third grade addresses the reading skills identified in the Common Core State Standards. According to these standards, third grade is a year of reading mastery wherein students should have acquired the requisite foundational reading skills (Common Core Standards Initiative, 2016). In third grade, students should be skilled in applying phonics and word analysis skills in decoding words. They should also be able to read with fluency and accuracy to support comprehension and should be able to read in order to acquire knowledge, expand their vocabulary, and construct meaning. The Common Core State Standards Initiative (2016) expects third graders to read different types of fiction and nonfiction such as poetry, short stories, scientific articles, graphs, and glossaries as independently as they possibly can with minimal assistance from adults.

Renaissance Learning (2014) asserted that STAR is a reliable and valid measure of reading skills because of evidence that it is aligned to curriculum standards at the state and national levels, including the Common Core State Standards. They explained that content is a crucial facet of test validity; content-related evidence of validity lies in the degree of correspondence, or alignment, between knowledge and skills measured by an assessment's test items and the knowledge and skills intended to be taught and learned in a given curriculum at a given grade level or levels. (Renaissance Learning, 2014, p. 22)

Nevertheless, they conducted a test of STAR Reading's internal consistency and retest correlation coefficients in a random sample of over 1.2 million tests administered from September 2012 to June 2013 (Renaissance Learning, 2014). Internal consistency,

according to Vogt (2005), is the extent to which items are correlated to one another. Retest reliability, on the other hand, is the “coefficient of correlation between pairs of test scores earned by the same students on different occasions” (Renaissance Learning, 2014, p. 19). According to Hopkins (2000), a correlation of 1.0 represents perfect agreement between tests, and a correlation of 0 indicates a complete lack of agreement. George and Mallery (2003) offered the following rule of thumb when interpreting reliability results: 0.9 = Excellent, 0.8 = Good, 0.7 = Acceptable, 0.6 = Questionable, 0.5 = Poor, and 0.4 and below = Unacceptable. The generic reliability internal consistency of STAR reading was calculated at $r = 0.97$ for the test overall, while it was calculated at $r = 0.94$ for the third grade test. The retest reliability was reported at $r = 0.9$ for the overall STAR Reading test and $r = 0.75$ for the third grade Reading test; both coefficients represent a high reliability.

Renaissance Learning (2014) also conducted predictive and concurrent validity tests of STAR Reading. Predictive validity, according to Vogt (2005, p. 244), is the “extent to which a test, scale, or other measurement predicts subsequent performance or behavior.” Renaissance Learning (2014) reported that the predictive validity of STAR Reading in third grade is $r = .80$, a high predictive validity. Concurrent validity, on the other hand, is a “way of determining the validity of a measure by seeing how well it correlates with some other measure the researcher believes is valid” (Vogt, 2005, p. 54). STAR Reading’s concurrent validity for third grade, according to Renaissance Learning (2014), is $r = .75$, which is a strong concurrent validity.

Results from the STAR Reading assessment are made available to the teacher immediately after the student finishes the test. The STAR Reading scores provided for

this particular study are norm-referenced scores, specifically percentile rank, which compare a student’s test results to the results of other students who have taken the same test (Renaissance Learning, 2012). STAR Reading Percentile Rank scores range from 1 to 99 and express student ability relative to the scores of other students in the same grade (Renaissance Learning, 2012). For a particular student, the Percentile Rank score indicates the percentage of students in the norms group who obtained lower scores. Students who score at or above the 40th percentile are considered to be at or above standard in reading. Students in the 25th to the 39th percentile are deemed to be “on watch.” Those who are in the 10th to the 24th percentile are in need of intervention, while students in the 9th percentile or below are considered to require urgent intervention. Table 2 displays the STAR Reading Percentile Rank scores and range for third grade.

Table 2

Third Grade STAR Reading Percentile Rank Range

Percentile	Scaled Score – Fall	Scaled Score – Winter	Scaled Score – Spring
10	177	215	255
20	235	272	311
25	258	294	334
40	319	357	393
50	357	392	436
75	461	500	547
90	561	613	673

The developers of STAR Reading (Renaissance Learning, 2015b) claimed that STAR Reading is correlated with state standardized tests in the state of Washington,

California, Oregon, and Connecticut. They analyzed the STAR Reading scores and state standardized scores from 2014-2015 of over 50,000 students in eight large school districts in the above-mentioned states and reported strong Pearson correlations that ranged from .81 to .83 in third grade through eighth grade (Renaissance Learning, 2015b). However, statistical significance levels were not reported. Also, no information is provided relating to the demographic make-up of the school districts and the students whose scores were analyzed. Thus, these results must be interpreted with caution.

Smarter Balanced Assessment (English Language Arts/Literacy). The Smarter Balanced Assessment (SBA) English Language Arts/Literacy is an untimed, summative state exam administered to students in the state of Washington in grades third through eighth, tenth, and eleventh. The SBA includes multiple choice questions, constructed-response items, and performance tasks. According to the Washington State Office of the Superintendent of Public Instruction (OSPI, 2016), the performance tasks are designed to evaluate skills such as depth of understanding, research skills, and complex analysis, which cannot be adequately assessed with selected- or constructed-response items. Open-ended questions are scored by professional scorers who undergo stringent training, and the validity and reliability of scoring are constantly monitored throughout the scoring process through double-scoring and read-behinds by scoring supervisors (Washington State Office of Superintendent of Public Instruction, 2016).

According to the Smarter Balanced Assessment Consortium (2015), the SBA reports scale scores, which are the students' overall numerical score. These scores fall on a continuous scale (from approximately 2000 to 3000). Scores that range from 2114 to 2366 are considered to be Level 1 (Minimal understanding of/ability to apply skills),

while Level 2 scores (Partial understanding of/ability to apply skills) range from 2367 to 2431. Level 3 (Adequate understanding of/ability to apply skills) is from 2432 to 2489, while Level 4 (Thorough understanding of/ability to apply skills) is from 2490 to 2632.

The Smarter Balanced Assessment in English Language Arts tests students in four areas or “claims”: reading, writing, speaking and listening, and research and inquiry. The overarching Common Core State Standards measured by these four claims are noted below (Smarter Balanced Consortium, 2015).

Claim 1 Reading. “Students can read closely and analytically to comprehend a range of increasingly complex literary and informational texts.”

Claim 2 Writing. “Students can produce effective writing for a range of purposes and audiences.”

Claim 3 Speaking and Listening. “Students can employ effective speaking and listening skills for a range of purposes and audiences.”

Claim 4 Research/Inquiry. “Students can engage in research and inquiry to investigate topics, and to analyze, integrate, and present information.”

Claim 1, which is the reading portion of the SBA English Language Arts/Literacy, measures students’ literary and informational reading skills according to the Common Core State Standards (Washington State Office of the Superintendent, 2016). Specifically, the reading claim of the SBA includes questions related to key details, central ideas, word meaning, reasoning and evidence, analysis within and across texts, text structures and features, and language use. Appendix B shows specific components of the reading claim of the SBA and the corresponding Common Core State Standard in third grade.

While the overall Smarter Balanced Assessment reports students' scale scores, the reading portion of the SBA reports only the students' reading level and the corresponding raw score. Level 1 means below standard, Level 2 means near or at standard, and Level 3 is above standard (Washington State Office of the Superintendent of Public Instruction, 2016). There is ambiguity around the Level 2 distinction because of a lack of guidance as to what constitutes "near standard" compared to "at standard." For statistical procedure purposes, it is challenging to examine the relationship between the STAR Reading test and the SBA Reading claim because of lack of clarity around the SBA Reading claim levels. Therefore, the researcher decided not to use the SBA Reading claim levels and instead used the overall SBA English Language Arts/Literacy test scores for this particular study. Even though the SBA ELA/Literacy test is a more global assessment than STAR Reading, it is logical to assume the two tests are not drastically different from one another because reading, to a certain extent, is woven throughout the entire SBA ELA/Literacy test. For instance, in order to complete the writing, listening and speaking, and research and inquiry components of the SBA, students must first read and understand a fairly lengthy amount of text giving them directions for the tasks and prompts. Furthermore, the research and inquiry claim as well as the writing claim require readings of articles, databases, and/or stories in order to complete the required tasks. The speaking and listening claim, on the other hand, requires students to tap into their vocabulary bank in order to understand the content and prompts and to produce the output required by the test. Mart (2012) found a relationship between reading and speaking skills, explaining that people who develop large vocabularies through reading also have large speaking vocabularies.

Procedure

STAR Reading is a progress monitoring measure first adopted in 2014-2015 by the school district participating in the current study. It was administered district-wide three times during the school year to students in second through ninth grades. For the purposes of this study, only third grade STAR Reading results were examined. Test administration in the participating school district occurred in individual classrooms in the fall, winter, and spring, with at least four weeks required between each test administration. STAR Reading results were made available to teachers and building administrators immediately following each test administration. While everyone was encouraged to administer the STAR Reading test, it was not a requirement of all teachers. Nevertheless, all third grade teachers in the district administered the STAR Reading assessment to their students.

The SBA, on the other hand, was a new state standardized exam in Math and English Language Arts that was administered statewide for the first time in 2014-2015. In the case of the school district participating in this study, the SBA was administered from April to June district-wide. Six elementary schools administered the test online, while three schools administered the test using the traditional paper-and-pencil method.

The researcher received oral and written permission from the school district superintendent to access preexisting 2014-2015 third grade STAR Reading and SBA data. The data provided to the researcher did not include any student identifiers such as student names. The Institutional Review Board at Seattle Pacific University granted approval for the use of archival data in March 2016.

Data Analysis

The research questions highlighted in this study required three different data analyses: Spearman's rank-order correlation, hierarchical multiple regression, and standard multiple regression.

Question 1. Is there a statistically significant relationship between the STAR Reading spring scores and the SBA English Language Arts/Literacy assessment scores in third grade?

Hypothesis 1. There is a statistically significant relationship between the STAR Reading spring scores and SBA English Language Arts/Literacy scores in third grade.

To answer this research question, the researcher calculated the Spearman's rank-order correlation, which measures the strength and direction of the association and relationship between two continuous or ordinal variables (Field, 2013). In this case, the STAR Reading spring score is an ordinal variable and the SBA English Language Arts/Literacy score is a continuous variable. Spearman's correlation coefficient, denoted by r_s and also called Spearman's rho, is a non-parametric statistic that is based on ranked data; it works by first ranking the data and then applying Pearson's equation to the ranked data (Field, 2013). The Spearman's rank-order correlation is appropriate because, for this particular question, the researcher is simply interested in looking at the association between the two assessment scores instead of making statements about causality or determining which variable causes the other to change.

There are no strict rules regarding the interpretation of the strength of relationship between two variables; however, Cohen (1988) provided a general rule of thumb.

According to Cohen (1988), the closer the value of the correlation coefficient is to zero,

the weaker the association between the two variables. Conversely, the closer the value of r_s is to ± 1 , the stronger the relationship is between the two variables. Thus, a coefficient value of ± 1.0 indicates a perfect relationship. A coefficient value of $\pm .05$ or higher typically means that two variables have a strong relationship, while $\pm .04$ to $\pm .03$ is deemed a moderate correlation. Below $\pm .03$ usually signifies a weak relationship, and a value of 0 indicates a nonexistent relationship between the variables. Based on literature and previous research conducted on the relationship between curriculum-based measure and high-stakes standardized assessments, the researcher hypothesized that there exists a relationship between the STAR Reading spring scores and the SBA English Language Arts/Literacy scores.

Question 2. Is there a statistically significant relationship between the STAR Reading assessment scores from fall, winter, and spring and the SBA English Language Arts/Literacy scores in third grade?

Hypothesis 2. There is a statistically significant relationship between the STAR Reading scores from fall, winter, and spring and the SBA English Language Arts/Literacy scores in third grade.

Hierarchical multiple regression was utilized to answer this research question. The main purpose of hierarchical multiple regression is to determine the proportion of the variation in the dependent variable explained by the addition of new independent variables (Field, 2013). In hierarchical regression, the predictor variables are entered into the model in a predetermined order instead of entering all of them at the same time, which is what is done in standard multiple regression (Field, 2013). Adding sets of variables, starting with the most important variable, in a predetermined order allows the

researcher to determine how much each set of variables uniquely adds to the prediction of the dependent variable (Field, 2013). For this particular research question, the STAR Reading spring scores, considered to be the most important variable because it was administered around the same time as the SBA, were added into the regression model first, followed by the STAR Reading winter and STAR Reading fall scores.

Question 3. Is there a statistically significant relationship between students' performance on SBA English Language Arts/Literacy and their STAR Reading spring scores, gender, Free and Reduced Lunch status, and Special Education status in third grade?

Hypothesis 3. There is a statistically significant relationship between students' SBA English Language Arts/Literacy scores and their STAR Reading spring scores, Free and Reduced Lunch status, and Special Education status in third grade.

The researcher was interested in exploring the relationship between the SBA ELA/Literacy scores and the students' gender, Free and Reduced Lunch status, and Special Education status, in addition to the STAR Reading spring score, because of research suggesting a reading achievement gap related to gender, low income status, and Special Education qualification. Student ethnicity is another factor research has shown to create a reading achievement gap, with many minority students not performing at standard in reading. However, for this particular study, student ethnicity was not added to the multiple regression model because the sample was made up of more than 75% white students. The remaining students comprised very small ethnicity subgroups, so that it would be a challenge to generate statistically significant results.

To answer the third research question, standard multiple regression was

conducted. Standard multiple regression, according to Field (2013), is an “extension of simple regression in which an outcome is predicted by a linear combination of two or more predictor variables” (p. 880). For this particular question, the SBA English Language Arts/Literacy score was the outcome or dependent variable, and the independent variables were STAR Reading spring scores, gender, socioeconomic status, and SPED status. Multiple regression determines the overall fit of the model (i.e., variance explained) and each independent variable’s contribution to the total variance explained. Using multiple regression allowed the researcher to determine the proportion of the variation in SBA English Language Arts/Literacy (outcome or dependent variable) that can be explained by the independent variables in the regression model: STAR Reading spring scores, gender, socioeconomic status, and SPED status.

The results of the data analyses for the three research questions and three hypotheses presented in this study are reported in Chapter Four. Interpretations of the data results are also included in that chapter.

Chapter Four

Results

This study was conducted to investigate the relationship between third grade students' scores in a progress monitoring measure in reading called STAR Reading and in the Washington State standardized exam, the Smarter Balanced Assessment in English Language Arts/Literacy.

The researcher analyzed archival 2014-2015 third grade reading data from a medium-sized, semirural school district in the Pacific Northwest to answer three research questions posed in this study. These questions are as follows.

1. Is there a statistically significant relationship between STAR Reading spring scores and the SBA English Language Arts/Literacy scores in third grade?
2. Is there a statistically significant relationship between the students' STAR Reading assessment scores from fall, winter, and spring and their SBA English Language Arts/Literacy scores in third grade?
3. Is there a statistically significant relationship between SBA English Language Arts/Literacy scores and STAR Reading spring scores, gender, Free and Reduced Lunch status, and SPED status?

The above-mentioned research questions warranted three different types of statistical procedures, namely: Spearman's rank-order correlation, hierarchical multiple regression, and standard multiple regression. The results of the data analyses are discussed in this chapter.

Data Results I

Research Question 1. Is there a statistically significant relationship between STAR Reading spring scores and the SBA English Language Arts/Literacy assessment scores in third grade?

Hypothesis 1. There is a statistically significant relationship between STAR Reading spring scores and SBA English Language Arts/Literacy scores in third grade.

The study measured the relationship between the STAR Reading spring scores and the SBA English Language Arts/Literacy scores because both tests were administered about the same time, in the spring of the 2014-2015 school year. Spearman's rank-order correlation was used to determine the relationship between the two variables. Spearman's correlation is similar to Pearson's correlation in that both procedures calculate the strength and direction of the association between two variables (Field, 2013). The difference, according to Field, is that Pearson correlation uses standard deviation, while Spearman's correlation examines differences in ranks of observations rather than the numeric values. Pearson correlation requires the use of two continuous variables (Field, 2013), and because one of the variables involved in this research question is ordinal (STAR Reading spring score is reported in percentile rank and is therefore an ordinal variable), the use of Pearson correlation is statistically inappropriate and may lead to incorrect or misleading results.

To be able to calculate Spearman's rank-order correlation coefficient or Spearman's rho, certain statistical assumptions about the characteristics of the data must have been met for interpretations of results to be as accurate as possible (McDonald, 2014). Violation of these assumptions may adversely influence research outcomes,

interpretations, and conclusions. The assumptions that must be met for Spearman's rho are as follows (McDonald, 2014).

Variables. The two variables must be measured either on a continuous and/or ordinal scale. That is, the two variables could be both continuous, both ordinal, or continuous and ordinal. The two variables considered for this particular research question were continuous (SBA English Language Arts/Literacy score) and ordinal (STAR Reading spring score). Therefore, this assumption has been met.

Paired observations. The two variables must represent paired observations. This research question involves two paired observations (SBA English Language Arts/Literacy scores and STAR Reading spring score). This assumption has, therefore, been met.

Monotonic relationship. The two variables must have a monotonic relationship. A monotonic relationship is a relationship that represents the following: (a) as the value of one variable increases, so does the value of the other variable, or (b) as the value of one variable increases, the value of the other variable decreases (McDonald, 2014). Visual inspection of a scatterplot with the variables SBA score and STAR Reading spring score shows a monotonic relationship; therefore, this assumption has not been violated.

Because Spearman's rank-order correlation is a non-parametric test, it does not require normality of distribution like Pearson correlation does.

With all three assumptions met, the Spearman correlation was an appropriate statistical method to use to answer the first research question. Results showed a strong, statistically significant, positive monotonic correlation between STAR Reading spring scores and SBA English Language Arts/Literacy scores ($r_s = .805, n = 557, p < .01$). A

correlation coefficient of .805, according to Cohen (1988), indicates a strong relationship between the two variables. This result confirms the hypothesis that there is a statistically significant relationship between the SBA English Language Arts/Literacy and STAR Reading spring scores.

One important consideration to note is that the original data set for this study included 651 students in the 2014-2015 third grade cohort in the participating school district. However, when running Spearman's correlation on the statistical analysis package, Statistical Package for the Social Sciences or SPSS, only 557 cases were included in the calculation because those with any missing values were deleted. Excluding missing data from the statistical procedure is typically a default procedure in statistical packages like SPSS (Briggs, Clark, Wolstenholme, & Clark, 2003). This method of excluding cases with missing values from the statistical procedure seems to be the most appropriate solution, given the goals of the study and the procedure's advantages and disadvantages relative to that of the other methods.

One major advantage of missing data exclusion is that it requires no special computational methods and can be used with any statistical procedure (Soley-Bori, 2013). However, Gelman and Hill (2007) warned that excluding data may result in estimates with larger standard errors due to reduced sample size. Field (2013) defined standard error as the "standard deviation of sample means. As such, it is a measure of how representative a sample is likely going to be of the population" (p. 54). Field (2013) explained that a "large standard error (relative to the sample means) means a lot of variability between the means of the different samples" (p. 54); thus, the sample might not be reflective of the population, yielding a less precise estimate of the population

means. A small standard error, on the contrary, signifies that the sample is likely to be representative of the population. Taking Gelman and Hill's (2007) word of caution into consideration, reduced sample size does not appear to be a concern in this particular instance because, even with missing data being excluded, the sample size for this research question remains fairly large ($N = 557$); a large sample size in regression generally being $N > 77$ (Field, 2013). Additionally, the standard error of the mean for the STAR Reading spring scores and the SBA ELA/Literacy score was fairly small, indicating that the sample means were likely to be an accurate representation of the population mean ($SE = 1.16$, $M = 59.54$ for STAR Reading spring score and $SE = 3.23$, $M = 2447.92$ for SBA ELA/Literacy). However, these results need to be interpreted with caution because this particular study, as a whole, used convenience sampling instead of random sampling. Thus, it is very likely that the sample is not fully representative of the population to begin with.

Gelman and Hill (2007) also contended that excluding cases with missing data can lead to biased estimates if the missing data are not considered Missing Completely at Random (MCAR). MCAR means the missing data are not associated with any of the variables, missing or observed (Gelman & Hill, 2007). In other words, there should be no pattern or system that makes some data more likely to be missing than others; the missing values should merely be a random subset of the complete data package. Hence, excluding missing values that are missing completely at random would yield the same result as the full data set would. Unfortunately, it is not possible to fully know if the missing values are missing completely at random (Gelman & Hill, 2007; Molenberghs et al., 2004). Little's Test for MCAR is a commonly used statistical test in an attempt to decipher if

missing values are completely at random, however, it is neither definitive nor completely accurate (Dong & Peng, 2013). The test is a chi-square statistic, the null hypothesis of which is that the data are missing completely at random. A statistically significant result ($p < .05$) fails to reject the null hypothesis and indicates the missing data are mostly likely not completely missing at random (Little, 1988). Results of the Little's MCAR test for this research question showed statistically significant results ($Chi-Square = 75.332, DF = 26, p = .000$). Because the test results were statistically significant, the results failed to reject the null hypothesis, which means the data are likely not missing completely at random, introducing potential bias to the results. Therefore, the statistical results of the Spearman's correlation need to be interpreted with caution.

Besides excluding cases with missing values in the statistical procedure, a common way to handle missing data is mean imputation or substitution. Mean imputation entails "filling in the missing values with a plausible one, such as the mean for the cases of the observed variable" (Pigott, 2001, p. 365). One disadvantage of this procedure, according to Pigott, is that it does not preserve the relationship among variables. This poses a major problem for this research question, which primarily aims to examine the relationship between the SBA English Language Arts/Literacy scores and the STAR Reading spring score. Therefore, the researcher decided not to use mean substitution to deal with missing data in this particular scenario.

There are other ways to deal with missing data, but they all present one issue or another. For instance, regression imputation, which replaces missing values with a predicted score from a regression equation, may overestimate the model fit and distort the correlation estimates (Durrant, 2005). Multiple imputation, which entails filling in

missing values with imputed values using a specified regression model, is very cumbersome and leaves much room for error (Horton & Kleinman, 2007). If the specification of the imputation model is not done appropriately, this can lead to potential bias in results (Horton & Kleinman, 2007).

Given no one perfect method to deal with missing data, the researcher carefully considered available options and decided to exclude cases with missing data in the statistical analysis because missing data exclusion seems to have the least adverse impact on the results of the Spearman's correlation.

Data Results II

Research Question 2. Is there a statistically significant relationship between the STAR Reading assessment scores from fall, winter, and spring and the SBA English Language Arts/Literacy scores in third grade?

Hypothesis 2. There is a statistically significant relationship between STAR Reading assessment scores from fall, winter, and spring and SBA English Language Arts/Literacy scores in third grade.

To answer this research question, hierarchical multiple regression was conducted, with the SBA English Language Arts/Literacy score as the dependent variable and the STAR Reading fall, winter, and spring scores as the independent variables. Hierarchical multiple regression is a statistical procedure very similar to standard multiple regression in that a linear combination of two or more predictor variables is used to predict an outcome (Field, 2013).

Much like standard multiple regression, the basic goal of hierarchical multiple regression, according to Field (2013), is to assess how much variance in a continuous

dependent variable can be explained by a set of independent variables. The difference between standard and hierarchical multiple regression is that hierarchical multiple regression allows the entry of the independent variables into the regression equation in a specific order as determined by the researcher and not by the computer, which is the case with standard multiple regression (Field, 2013). Hierarchical multiple regression, according to Petrocelli (2003), is useful to researchers who are

interested in testing theoretical assumptions and examining the influence of several predictor variables in a sequential way, such that the relative importance of a predictor may be judged on the basis of how much it adds to the prediction of a criterion, over and above that which can be accounted for by other important predictors. (p. 10)

Petrocelli (2003) added that the focus of hierarchical multiple regression is “on the change in predictability associated with predictor variables entered later in the analysis over and above that contributed by predictor variables entered earlier in the analysis” (p. 11). Therefore, the analysis outcome may be largely dependent on the order in which variables are entered into the equation. Field (2013) posited that as a general rule, predictors need to be “...entered into the model first in order of their importance in predicting the outcome” (p. 322). For this particular research question, hierarchical multiple regression was appropriate because the researcher was interested in investigating the additional importance of two independent variables (STAR Reading fall score and STAR Reading winter score) in predicting the dependent variable (SBA score) over and above another independent variable (STAR Reading spring score). The researcher maintains that logically the STAR Reading spring score is the most important predictor

variable because it was administered towards the end of the school year, closest to the time that the SBA was administered, and after the students received about a year's worth of instruction in reading and literacy. Thus, the STAR Reading spring score was entered into the regression model first. Furthermore, the researcher believes the STAR Reading fall and STAR Reading winter scores are also important predictors of SBA scores and that teachers should carefully analyze them in order to make adjustments to their instruction as the year progresses. Therefore, the STAR Reading fall and STAR Reading winter scores were entered next to see if they explain the variance in the SBA ELA/Literacy scores over and above the STAR Reading spring scores.

Prior to conducting a hierarchical multiple regression, a number of assumptions pertinent to this statistical procedure were tested. These assumptions needed to be tested in order to avoid biased or misleading results (Laerd Statistics, 2015; Berry, 1993). They are discussed below.

Continuous dependent variable. The study must have a continuous dependent variable. This study meets this assumption because the SBA English Language Arts/Literacy scores, reported as scaled scores, are a continuous dependent variable.

Two or more independent variables. The research design must have two or more independent variables, which can either be continuous (interval or ratio variable) or categorical (ordinal or nominal). This study meets this assumption because for this specific question, three independent variables are taken into consideration (STAR Reading fall scores, STAR Reading winter scores, and STAR Reading spring scores). The independent variables are all ordinal variables.

Independence of residuals. The study must have independence of residuals, which is defined as “the difference between the value a model predicts and the value observed in the data on which the model is based” (Field, 2013, p. 883). In other words, the residuals cannot be related to one another, or an alternative statistical procedure will need to be run. The Durbin-Watson statistic, which can range in value from 0.0 to 4.0, tests the serial correlation between residuals and was used to test this assumption (Field, 2013). The value of approximately 2.0 indicates that there is no correlation between residuals; the closer the value is to 2.0, the more likely that the residuals are independent of each other (Field, 2013). This study had a Durbin-Watson statistic of 1.97. Because the value of the Durbin-Watson statistic was very close to 2.0, it can be accepted that there is independence of residuals.

Linear relationship. There needs to be a linear relationship between the dependent variable and the independent variables taken together and separately. Appendix C includes scatterplots that show the linear relationship between the dependent variable and the independent variables taken collectively and separately.

Homoscedasticity of residuals. The data must show homoscedasticity of residuals, which means that the residuals at each level of the predictor(s) should have the same or equal variance (Field, 2013). This study has met this particular assumption as assessed by visual inspection of the scatterplot represented in Figure 1, which shows the residuals being randomly scattered and approximately constantly spread, exhibiting no pattern.

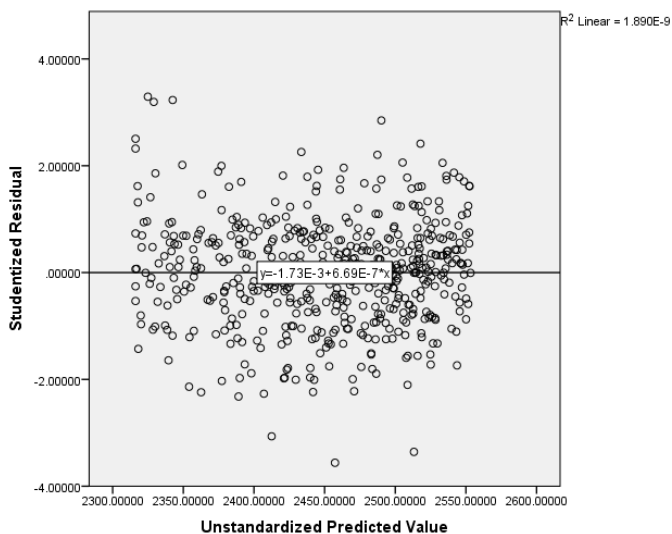


Figure 1. Scatterplot Showing Homoscedasticity

Multicollinearity. The data must not show multicollinearity. Multicollinearity occurs when two or more independent variables are highly correlated to one another (Field, 2013), which can present issues in understanding which independent variable contributes to the variance explained in the dependent variable. A Tolerance value of less than 0.1, after collinearity statistics on SPSS are run, indicates a serious collinearity problem, while values below 0.2 means that there may potentially be a collinearity issue (Field, 2013). Results of collinearity statistics for this specific question showed that the Tolerance values ranged from .233 to .248; there was, therefore, no reason to be concerned about the independent variables highly influencing one another.

Outliers. There should be no significant outliers, high leverage points, or highly influential points in the data. Outliers and leverage and influential points represent observations that deviate from the main trend of the data and may have adverse effects on the regression model (Field, 2013).

One way to detect outliers is to use casewise diagnostics. The casewise diagnostics table highlights any cases where the standardized residual is greater than ± 3

standard deviations (Field, 2013). A value of greater than ± 3 is used to define whether a particular residual might be representative of an outlier or not. For this particular study, only six cases are identified to have values larger than ± 3 standard deviations, with the most extreme case having a value of -3.56. However, none of these cases had a large leverage value or influence, as discussed below, so they were not removed from the data set.

Another measure of influential data is leverage, which “gauges the influence of the observed value of the outcome variable over the predicted values” (Field, 2013, p. 307). To determine whether there are any cases that exhibit high leverage or exert undue influence on the model, a general rule of thumb is to consider leverage values of less than 0.2 as safe. For this particular study, all cases had values that were less than 0.2 and were considered to be in the safe range.

The Cook’s Distance in linear regression is “a measure of the overall influence of a case on the model” (Field, 2013, p. 306). As a general rule of thumb, if there are Cook’s Distance values above 1.0, they should be examined as they may present problems according to Field (2013). For this particular research question, there were no Cook’s Distance values above 1.0.

Normality of distribution. The last assumption that needs to be met in order to run hierarchical regression is that residuals are approximately normally distributed, with a mean of zero (Field, 2013). This means that “the difference between the model and the observed data are most frequently zero or very close to zero, and that differences much greater than zero happen only occasionally” (Field, 2013, p. 311). One common method for checking this assumption is to inspect a histogram and a P-P plot or a normal Q-Q

plot of the studentized residuals (Laerd Statistics, 2015). A visual inspection of the histogram run for this study showed an approximately normal distribution of the residuals. Inspection of the P-P plot showed that the points are aligned well along the diagonal line, close enough to normal for the hierarchical regression analysis to proceed. A hierarchical multiple regression analysis is fairly robust against deviations from normality (Laerd Statistics, 2015), so these results can be accepted as not violating the assumption of normality. Appendix D includes the histogram and the P-P plot.

Based on the above discussion, the data in this study met all the assumptions of hierarchical multiple regression. Hierarchical regression was then run on SPSS to determine if the addition of STAR Reading winter and STAR Reading spring scores explained the variance in the dependent variable over and above STAR Reading fall scores alone. Table 3 displays the full details on each regression model.

Table 3

Hierarchical Regression Model Summary

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. Error of the Estimate	Change Statistics				
					ΔR^2	ΔF	<i>df</i> 1	<i>df</i> 2	<i>Sig.</i> ΔF
1	.789 ^a	.623	.622	48.13	.623	872.07	1	528	.000
2	.831 ^b	.690	.689	43.69	.067	57.33	2	526	.000

a. Predictors: (Constant), STAR Spring Score

b. Predictors: (Constant), STAR Spring, STAR Winter, STAR Fall

As shown in Table 3, the full model of STAR Reading fall, winter, and spring scores (Model 2) has a statistically significant relationship to the third grade students' performance on the SBA English Language Arts/Literacy ($R = .831$). *R* is the multiple correlation coefficient, which measures the strength of the association between the

independent variables (STAR Reading fall scores, STAR Reading winter scores, and STAR Reading spring scores) and the dependent variable (SBA ELA/Literacy scores). The closer the multiple correlation coefficient is to 1.0, the stronger the relationship (Field, 2013). An R of zero means that there is no linear relationship between the independent variables and the dependent variables. Unlike the Pearson's correlation coefficient, which indicates both the strength and direction of the relationship, the multiple correlation coefficient only tells the strength of the association. For this research question, Model 2, which includes all three independent variables, shows a strong level of association ($R = .831$) between the dependent and independent variables. The statistical significance of the overall model as assessed by the analysis of variance (ANOVA) was $p < .001$. This confirms the hypothesis for this research question.

The value of R^2 or coefficient of determination as displayed in Table 3 is a measure of how much of the variability in the outcome is explained by the independent variables. The results presented in Table 3 showed that the second model explains a greater amount of variation in the dependent variables as more variables were introduced ($R^2 = .623$ for Model 1 and $.690$ for Model 2). That is to say that the models got slightly better at explaining the variance in the dependent variable with the addition of STAR Reading fall and STAR Reading winter scores. In the first model, R^2 is equal to $.623$, which means that the STAR Reading spring score accounted for 62.3% of the variation in the SBA scores. When the STAR Reading fall and winter scores were added (Model 2), R^2 increased slightly to $.690$. This means that the entire model (STAR Reading spring score and the added variables of STAR Reading fall and winter scores) accounted for 69.0% of the variability in the SBA ELA/Literacy score. This is a fairly small but

statistically significant ($p < .001$) increase in R^2 of .067 that can be attributed to the addition of the STAR Reading fall and winter scores to the model. In other words, STAR Reading fall and winter scores accounted for 6.7% of the variance in the SBA scores over and above the STAR Reading spring scores. In short, STAR Reading fall and STAR Reading winter scores had a small but statistically significant contribution to explaining the variance in the outcome variable (SBA English Language Arts/Literacy) above and beyond STAR Reading spring scores.

The coefficient of determination or R^2 is based on the sample; therefore, it is considered to be a positively biased estimate of the proportion of the variance of the dependent variable explained by the regression model (Field, 2013). This means that R^2 is bigger than it should be when generalizing to a larger population. The adjusted R^2 is a measure that corrects for this positive bias because it represents the coefficient of determination that one would expect in the larger population (Field, 2013). The value of the adjusted R^2 for Model 2 is .689. This indicates that the inclusion of the independent variables into the regression model explained 68.9% of the variability of the dependent variable. It must be noted, however, that the adjusted R^2 is not that much smaller than the value of R^2 . Adjusted R^2 is also an estimate of effect size (Field, 2013), so a value of .689 is indicative of a large effect size according to Cohen's (1988) classification. In sum, the R^2 for the overall model was .690, with an adjusted R^2 of .689, which is a large size effect according to Cohen (1988). However, this result needs to be interpreted with caution because convenience sampling was used for this study, which means that the sample may not be fully representative of the population like random sampling would.

The Hierarchical Multiple Regression Coefficients Table, shown in Table 4, contains b -values that reflect the extent to which each of the variables in the regression model is associated with the dependent variable (SBA English Language Arts/Literacy) if the effects of all the other variables were held constant (Field, 2013). The unstandardized coefficient, B , reflects the change in the SBA ELA/Literacy scores (dependent variable) for every one unit change in the STAR spring scores (independent variable) if all other variables were held constant. The standardized b -values, β , on the other hand, show the number of standardized deviations that the outcome will change as a result of one standard deviation change in the predictor (Field, 2013, p. 140). Because the standardized coefficients are all expressed in terms of standard deviation, they are directly comparable to one another and better provide insight into each variable's contribution compared to the other variables. Field (2013) wrote that the bigger the absolute value of the standardized b -value, the more important the variable. According to Table 4, the STAR Reading spring scores ($\beta = .34$) is the variable in Model 2 that has the strongest relationship to the SBA scores. If all other variables were held constant, as STAR Reading spring scores increased by one standard deviation, the SBA ELA/Literacy scores also increased by .34 standard deviation. The standard deviation for SBA scores is 78.30 according to the descriptive statistics that were run, so a one standard deviation increase of the STAR spring scores would constitute a change of 26.62 points (.34 x 78.30) on the SBA ELA/Literacy scores. Looking at the standardized b -values in Table 4, STAR Reading fall scores are only slightly less important than STAR Reading spring scores and slightly more important than STAR Reading winter scores.

Table 4

Hierarchical Multiple Regression Coefficients

Variable	Model 1				Model 2			
	<i>B</i>	β	<i>t</i>	<i>Sig.</i>	<i>B</i>	β	<i>t</i>	<i>Sig.</i>
Constant	2312.14			.000	2312.73		486.51	.000
STAR Spring	.98	.79	29.53	.000	.98	.34	6.96	.000
STAR Winter					.70	.26	5.09	.000
STAR Fall					.75	.29	5.86	.000

Note: Significance level of all variables is $p < .05$. Dependent variable: SBA Score

The *t*-test statistic, another measure of whether the variable is making a significant contribution to the model, is associated with the *b*-values (Field, 2013). Field noted that if the *t*-test statistic is statistically significant, then the *b*-value is also statistically significant. The smaller the *p*-value and the larger the value of *t*, the larger the contribution of that predictor to the model. In Model 2, all three variables have a statistical significance value of $p < .001$. The STAR Reading spring score has the biggest *t*-value of 6.96, which means that of the three test administrations, it is the most important.

To summarize, a hierarchical multiple regression was run to determine if the addition of STAR Reading fall and STAR Reading winter improved the prediction of SBA ELA/Literacy over and above STAR Reading spring alone. There was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistics of 1.970. There was no multicollinearity, as assessed by Tolerance values

greater than 0.1. There were only six studentized deleted residuals greater than ± 3 standard deviations. There were no leverage values greater than 0.2 and no values for Cook's distance above 1.0. The assumption of normality was met, as assessed by visual inspection of a histogram and a P-P plot. The full model of STAR Reading fall, STAR Reading winter, and STAR Reading spring scores predicting the SBA ELA/Literacy score was statistically significant ($R^2 = .69$, $F(3, 526) = 390.93$, $p = .000$).

As is the case with the first research question, there are missing data involved in the second research question. Specifically, of the 651 cases or participants in the full data set, only 530 were included in the statistical analysis. SPSS defaults to excluding any cases with values missing in any of the variables when running hierarchical multiple regression. As with the Spearman rank-order correlation analysis discussed in the first research question, reduction in sample size, as a result of missing data exclusion, was not a huge concern for this particular research question. A sample size of 530 is still fairly large. The standard error of the mean for all the variables is fairly small, which indicates that the sample means of the variables are a fairly precise estimate of the population mean. However, this result needs to be interpreted with caution because convenience sampling, not random sampling, was used in the study. Table 5 shows the standard error of the mean for each of the variables with missing values.

Gelman and Hill (2007) cautioned that excluding all cases with missing data, also called listwise deletion or complete case analysis, may result in biased estimates if the excluded cases are not deemed missing completely at random (MCAR). Little's Test for MCAR, a commonly used but not definitive nor completely accurate test to determine if missing values are completely at random, was conducted for this research question.

Results showed statistically significant results ($Chi-Square = 75.33$, $DF = 26$, $p = .000$). Because the test was statistically significant, the results did not reject the null hypothesis, which means that the data are likely not missing completely at random. This could introduce potential bias to the outcome; therefore, interpreting the statistical results of the hierarchical regression with caution is important.

Table 5

Frequency Statistics

		SBA Score	STAR Reading Fall	STAR Reading Winter	STAR Reading Spring
<i>N</i>	Valid	594	571	585	577
	Missing	57	80	66	74
Mean		2447.93	50.13	56.15	59.54
Std. Error of Mean		3.23	1.27	1.20	1.16

There are alternative ways to deal with missing data as discussed previously in the first research question. However, they too have their sets of limitations and risks that may lead to potential bias in results. Mean substitution is not appropriate to use in handling missing data for this research question because it fails to preserve the relationship among variables (Pigott, 2001). This presents a major challenge for this kind of study that seeks to examine relationships between the dependent variable and the independent variables. Regression imputation, which replaces the missing values with a predicted score from a regression equation, may overestimate the model fit, so it does present its own set of challenges (Durrant, 2005). Multiple imputation, on the other hand, can lead to potential bias if the imputation models are erroneously specified (Horton & Kleinman, 2007).

Given this, the researcher decided to use listwise deletion, the exclusion of all cases with missing data, because its disadvantages are relatively less impactful on the results of the regression analysis compared to the alternatives.

Data Results III

Research Question 3. Is there a statistically significant relationship between students' performance on SBA English Language Arts/Literacy and their STAR Reading spring scores, gender, Free and Reduced Lunch status, and SPED status in third grade?

Hypothesis 3. There is a statistically significant relationship between students' SBA English Language Arts/Literacy scores and their STAR Reading spring scores, gender, Free and Reduced Lunch status, and SPED status in third grade.

Multiple regression, which is a statistical procedure that entails the use of a linear combination of two or more predictor variables to predict an outcome (Field, 2013), was conducted in order to answer this research question. For this research question, the predictor variables are gender, Free and Reduced Lunch status, SPED status, and STAR spring scores. The outcome or dependent variable is SBA ELA/Literacy. These predictor variables were chosen because previous research and literature have pointed to an achievement gap in reading between boys and girls, between low-income and non-low-income families, and between SPED and non-SPED students. The researcher was thus interested in exploring if these variables, in addition to STAR Reading spring scores, explain the variance in third grade SBA ELA/Literacy scores. Ethnicity is another variable identified by research and literature as an important factor in a student's reading achievement. However, it is not included in the regression model for this particular study

because the participants were primarily Caucasian students, and the ethnicity subgroups did not have a large enough sample size to yield statistically significant results.

Before running the standard multiple regression, eight assumptions needed to be tested and met to avoid the risks of misleading and biased results (Field, 2013). These assumptions are discussed below.

Continuous dependent variable. The research design must have a continuous dependent variable. This specific research question meets this assumption because the SBA English Language Arts/Literacy scale scores are a continuous dependent variable.

Two or more independent variables. The study needs to have two or more independent variables, which can either be continuous (interval or ratio variable) or categorical (ordinal or nominal). This assumption is met because for this particular study, multiple independent variables are included in the regression model (STAR Reading spring scores, gender, Free and Reduced Lunch status, and Special Education status). These independent variables are continuous or categorical variables.

Independence of residuals. The research design must have independence of residuals. This study has met this assumption, as assessed by a Durbin-Watson statistic of 1.91. The Durbin-Watson value of approximately 2.0 indicates that there is likely no correlation between residuals (Field, 2013).

Linear relationship. There needs to be a linear relationship between the dependent variable and the independent variables taken collectively and separately. A visual inspection of various scatterplots (see Appendix E) revealed that there is a linear relationship between the dependent variable and the independent variables taken all together and taken separately.

Homoscedasticity of residuals. The data must demonstrate homoscedasticity of residuals or equal error variances. The data in this study showed homoscedasticity, as assessed by a visual inspection of the scatterplot in Figure 2. The scatterplot in Figure 2 shows the residuals showing no clear pattern and instead being randomly scattered and spread out.

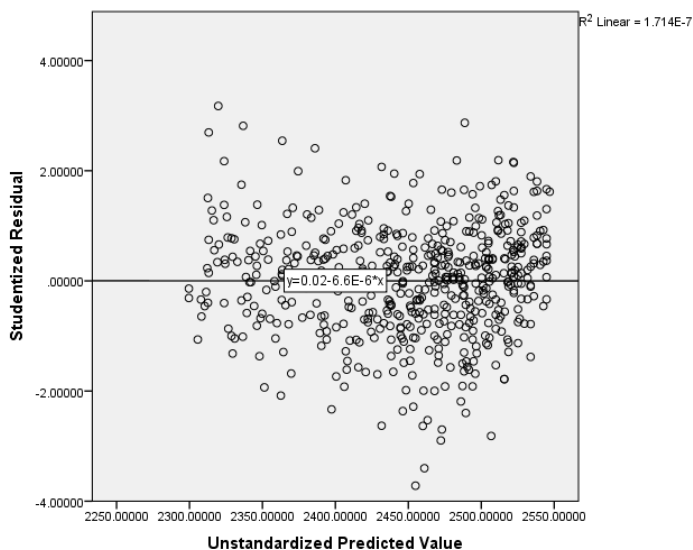


Figure 2. Multiple Regression Scatterplot Showing Homoscedasticity of Residuals

Multicollinearity. The data must not show multicollinearity, which happens when two or more independent variables are highly correlated to one another, making it difficult to gauge the individual importance of a predictor variable (Field, 2013). There was no evidence of multicollinearity in this particular research question, as assessed by Tolerance values greater than 0.1. All Tolerance values for this research question range from .832 to .981.

Outliers. There should be no significant outliers, high leverage points, or highly influential points in the data set because they can have an adverse effect on the regression equation. According to Casewise Diagnostics run for this particular study, only three

cases were identified as outliers as they have values larger than ± 3 standard deviations. The most extreme case had a value of -3.737. Because none of these cases had a large leverage value or influence, as discussed below, they were not removed from the data set.

In order to ascertain if there are any cases that demonstrate high leverage, a general rule of thumb for interpreting leverage values is that values less than 0.2 are generally safe (Field, 2013). For this particular research question, all cases had values less than 0.2.

The Cook's Distance in linear regression measures the influence that a case has on the on the regression model as a whole or "the impact that a case has on the model's ability to predict all cases" (Field, 2013, p. 306). As a general rule of thumb, Cook's Distance values above 1.0 present potential issues that should be examined for undue influence. For this particular study, there were no Cook's Distance values above 1.0.

Normality of distribution. The research design must have residuals, or errors in prediction, that are approximately normally distributed. Two common methods used to check for the assumption of normality of the residuals is checking a histogram with a superimposed normal curve and a P-P plot or a normal Q-Q plot of the studentized residuals (Laerd Statistics, 2015). A visual inspection of the histogram on Appendix G showed an approximately normal distribution of the residuals. The P-P plot included in Appendix G indicated that the points are aligned well along the diagonal line, close enough to normal for the standard multiple regression analysis to be conducted.

With all of the above mentioned assumptions met, it was appropriate to run standard multiple regression in order to determine the proportion of the variation in the SBA ELA/Literacy scores explained by the independent variables (STAR Reading spring

scores, gender, Free and Reduced Lunch status, and SPED status) and determine how much the dependent variable changes for a unit change in the independent variable. A multiple regression model summary can be found in Table 6 below.

Table 6

Multiple Regression Model Summary

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. Error of the Estimate
1	.801 ^a	.641	.639	47.15

a. Predictors: (Constant), SPED, Gender, FRL, STAR Reading Spring Score; $p < .001$

The multiple correlation coefficient, *R*, is simply a measure of the strength of the linear association between the dependent variable and the independent variable, with a value near 1.0 indicating a strong association (Field, 2013). For this particular research question, the multiple correlation coefficient is .801 (statistically significant at $p < .001$) as shown in Table 6, which indicates a strong linear relationship between the predictor variables (STAR Reading spring score, gender, Free and Reduced Lunch status, and SPED status) and the dependent variable (SBA ELA/Literacy score). This confirms the hypothesis that there is a statistically significant relationship between the SBA ELA/Literacy scores and the STAR Reading spring scores, gender, Free and Reduced Lunch status, and SPED status.

In order to assess if the multiple regression model is a good fit for the data, it is important to consider the value of the coefficient of determination, or *R*² as it is more commonly known. *R*² is a measure of the proportion of variance in the dependent variable that is explained by the independent variables (Field, 2013). In this study, the value of *R*² is equal to .641 ($p < .001$) as shown in Table 6, which means that the

inclusion of the independent variables (gender, Free and Reduced Lunch status, SPED status, and STAR Reading spring Score) into a regression model explained 64.1% of the variability of the dependent variable.

The value of the adjusted R^2 , which corrects for the positive bias of R^2 , for this particular research question is .639, which is only slightly smaller than the value of R^2 (.641). This means that the addition of all independent variables into the regression model explained 63.9% of the variability of the dependent variable, which is the value that one would expect in the population instead of the sample size. Adjusted R^2 is also an estimate of effect size, which at .639 is indicative of a large effect size according to Cohen (1988). In sum, the R^2 for the overall model was .641, with an adjusted R^2 of .639, which is a large size effect according to Cohen. However, it is important to note that the study's convenience sampling may not be representative of the state's student population, so these findings should be interpreted with caution.

The b -values displayed in Table 7 depict the relationship that each independent variable has to the dependent variable (Field, 2013). Moreover, the b -values, according to Field, represent "...to what degree each predictor affects the outcomes if the effects of all other predictors are held constant" (p. 338). The t -values shown in Table 7 are measures of whether the independent variable is making a significant contribution to the model (Field, 2013). If the t -test that is associated with the b -value is significant at $p < .05$, then the predictor variable has a statistically significant contribution to the model. The smaller the value of *Sig.* and the larger the value of t , the greater the contribution of that predictor. According to Table 7, STAR Reading spring score has a p -value of $< .001$ and the largest t -test value among all the independent variables ($t = 27.04$). This means that it

has the greatest contribution to the regression model compared to gender and Free and Reduced Lunch status. SPED status has no statistically significant contribution to the model.

Table 7

Multiple Regression Coefficients

Variable	<i>B</i>	<i>SE_B</i>	β	<i>t</i>	<i>Sig.</i>
Constant	2309.23	6.17		374.34	.000
STAR Spring Score*	2.17	.080	.76	27.04	.000
Gender*	22.42	4.04	.14	5.56	.000
FRL*	-11.87	5.11	-.06	-2.32	.021
SPED**	1.77	6.07	.01	.291	.771

Note. * $p < .05$; ** $p > .05$

As shown in Table 7, the unstandardized coefficient for STAR spring score is $B = 2.17$, significant at $p < .001$. The unstandardized coefficient reflects the change in the SBA ELA/Literacy score (dependent variable) for a one unit change in the STAR spring score (independent variable) when all other variables are held constant. As such, when all other variables are held constant, an increase in the STAR Reading spring score of one unit or percentile rank is associated with an increase in the SBA English Language Arts/Literacy score of 2.17 units or points, with a statistical significance level of $p < .001$. The multiple regression predicts that the higher the STAR Reading spring score, the higher the SBA English Language Arts/Literacy score.

A dichotomous variable such as gender as shown in Table 7 requires a different interpretation of the slope coefficient than that of a continuous variable. With a dichotomous independent variable, the value of the slope coefficient represents the

difference in the dependent variable between the two categories of the dichotomous independent variable, with the comparison based on the value of 0 (in this case males is the category assigned the value of 0 and females the value of 1) (Laerd Statistics, 2015). In other words, the coefficient represents the difference in predicted SBA English Language Arts/Literacy scores of females ($n = 274$) compared to males ($n = 255$). The unstandardized coefficient for this variable is 22.42 ($p < .001$), which means that, with all other things being equal, females have, on average, SBA English Language Arts/Literacy scores that are 22.42 points higher than males. This is consistent with previous research and literature regarding the gender achievement gap in reading.

Free and Reduced Lunch status is another dichotomous variable, with the value of 0 representing students not qualifying for free and reduced lunch ($n = 421$), and the value of 1 representing students that did qualify ($n = 108$). The unstandardized coefficient for this variable is -11.87 ($p < .05$), as shown in Table 7. This means that students who qualified for Free and Reduced Lunch scored, on average, 11.87 points lower on the SBA compared to their peers who did not qualify for Free and Reduced Lunch. This finding is also consistent with research and literature that have been published germane to the achievement gap between low-income and non-low-income students.

Special Education (SPED) Status is also a dichotomous variable included in the regression model. The value of 0 represents the students who did not qualify for SPED services, while the value of 1 represents those who qualified for SPED. However, this variable's contribution to the model is not statistically significant ($p = .771$). The statistically insignificant results are likely due to the small number of students in the

sample who qualified for Special Education ($n = 74$) in relation to the entire sample ($n = 651$).

The standardized b -values included in Table 7 offer another way to interpret the degree to which each variable affects the dependent variable. Unlike the unstandardized coefficient, the standardized b -values are not dependent on the units of the variable (Field, 2013). The standardized b -values represent the number of standardized deviations that the outcome will change as a result of one standard deviation change in the predictor (Field, 2013, p. 340). They provide an easier way to interpret the data, providing a better insight into the importance of each variable. According to Field (2013), the larger the absolute value of the standardized coefficient, the more important the variable. In this particular case, the STAR Reading spring score is the most important variable in the regression model, followed by gender and Free and Reduced Lunch status. Special Education status does not have a statistically significant contribution to the dependent variable.

In sum, a multiple regression was run to predict SBA ELA/Literacy scores from STAR Reading spring scores, gender, Free and Reduced Lunch status, and Special Education status. There was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistics of 1.908. There was homoscedasticity, as assessed by Tolerance values greater than 0.1. There were only three studentized deleted residuals greater than ± 3 standard deviations, no leverage values greater than 0.2, and no values for Cook's distance above 1. The assumption of normality was met, as assessed by visual inspection of a histogram and a P-P plot. The multiple regression model of STAR

Reading spring score, gender, Free and Reduced Lunch status, and Special Education status statistically significantly explained the variance in the SBA ELA/Literacy scores ($R^2 = .641$, $F(4, 552)$, 246.79 , $p = .000$, adjusted $R^2 = .639$).

As is the case with the first and second research questions, there are missing data that were excluded in the statistical analysis for this research question. The original data set had 651 cases. After listwise deletion or the exclusion of all cases with missing data, which was automatically run by SPSS, 557 cases were included. Similar to the first two research questions, a reduction in sample size that could lead to large standard error of the mean was not a significant concern for this particular research question, given that 557 is still a large sample size. The standard error of the mean for both the STAR Reading spring scores and the SBA ELA/Literacy scores is fairly small, which typically indicates that the sample is a good representation of the population. However, as with the first two research questions, this result should be interpreted with caution because this study used convenience sampling, instead of random sampling, which could minimize the chances of the sample representing the population to begin with. Table 8 shows the mean and standard error for the SBA score and the STAR Reading spring score with missing values.

According to Gelman and Hill (2007), listwise deletion or complete case analysis may generate biased estimates if the excluded cases are not determined to be missing completely at random (MCAR). Results of Little's Test for MCAR revealed statistically significant results ($Chi-Square = 22.16$, $DF = 2$, $p = .000$), which means that the results did not reject the null hypothesis and that the missing data were not missing completely at random. Because the missing data were not completely at random, the outcomes may

be biased and should be interpreted with caution. Nevertheless, the researcher decided to use listwise deletion to handle missing data, as opposed to mean imputation, regression imputation, multiple imputation, or other alternative methods, because it appeared to have the least adverse impact on the regression results compared to alternative procedures.

Table 8

Frequency Statistics

		SBA Score	Gender	FRL	SPED	STAR Reading Spring Score
<i>N</i>	Valid	594	651	651	651	577
	Missing	57	0	0	0	74
Mean		2447.93				59.54
Std. Error of Mean		3.23				1.16

Summary

There is a strong, positively monotonic, statistically significant relationship between the third grade SBA English Language Arts/Literacy scores and the STAR Reading spring scores. The hierarchical regression model that included STAR Reading fall scores, STAR Reading winter scores, and STAR Reading spring scores had a statistically significant relationship to the third grade students' SBA English Language Arts/Literacy scores. STAR Reading fall and STAR Reading winter scores had a small but statistically significant contribution to the variance in the SBA English Language Arts/Literacy scores over and above the STAR Reading spring scores. The multiple regression model that included gender, Free and Reduced Lunch Status, Special

Education status, and STAR Reading spring scores statistically significantly predicted the SBA English Language Arts/Literacy scores of the participating third grade students.

Chapter Five

Discussion

This chapter contains a summary and a discussion of the three hypotheses related to the research questions presented in this study. Connections between the study's findings and existing research and literature germane to reading and literacy are also discussed. Moreover, study limitations as well as suggestions for future research are also included in this chapter.

Summary of Findings

This research study was conducted to examine the relationship between third grade student performance on a reading progress monitoring assessment (STAR Reading) and on the Washington State standardized exam (Smarter Balanced Assessment in English Language Arts/Literacy). Specifically, Spearman's rank-order correlation was conducted to determine the relationship between the third grade STAR Reading scores from the spring test administration and the SBA ELA/Literacy scores. Results showed a statistically significant relationship between the two variables. A hierarchical multiple regression was run to investigate the relationship between the third grade SBA ELA/Literacy scores and the STAR Reading scores from the fall, winter, and spring test administration periods. Findings revealed that there was a statistically significant relationship between the three independent variables and the SBA ELA/Literacy scores. Secondly, hierarchical multiple regression was calculated to determine if the STAR Reading fall and STAR Reading winter scores added to the prediction of the SBA ELA/Literacy scores over and above the STAR Reading spring scores. According to the findings, both the STAR Reading fall and STAR Reading winter scores added to the

prediction of the SBA ELA/Literacy scores over and above the STAR Reading spring scores. Finally, standard multiple regression was conducted to explore the relationship between the dependent variable (third grade SBA ELA/Literacy scores) and the independent variables: gender, Free and Reduced Lunch status, Special Education status, and third grade STAR Reading spring scores. Results from the multiple regression analysis helped determine how much of the variance in the SBA ELA/Literacy scores was explained collectively and uniquely by the four independent variables included in the multiple regression model.

Discussion

Research Question 1. Is there a statistically significant relationship between students' STAR Reading spring scores and their SBA English Language Arts/Literacy assessment scores in third grade?

For research question 1, the researcher hypothesized that there is a statistically significant relationship between the participating students' STAR Reading spring scores and their SBA English Language Arts/Literacy scores. Results of the Spearman's rank-order correlation confirmed this hypothesis, revealing a statistically significant, strong, positively monotonic relationship between both variables ($r_s = .805, p < .01$). A positive, monotonic relationship in non-parametric statistics means that as the value of one of the variables increases, the value of the other variable also increases. According to this study, the third grade students' STAR Reading spring scores and their SBA ELA/Literacy scores were positively correlated at a statistically significant level so that as their scores on the STAR Reading spring test increased, so did their SBA ELA/Literacy scores.

The results of this study are instructive to schools and school districts who may be interested in gauging the feasibility and utility of a progress monitoring measure and determining whether it is worth the time, resources, and energy invested into it. One of the purposes of a progress monitoring measure is to track students' academic growth rate and proficiency level, according to set academic goals and learning standards, so that students may receive data-informed and appropriate intervention and enrichment. Findings from the Spearman's rank-order correlation analysis may impart insight into how well STAR Reading meets such an assessment objective. Educational leaders may view these findings as a validation of their existing efforts around the administration and use of STAR Reading as a progress monitoring measure or as a reason to firm up test administration practices and professional development in order to ensure that teachers understand the value of the test and know how to administer it and analyze the data. They may also find these results constructive as they put in place intervention systems designed to address reading gaps and deficiencies in students. Teachers, on the other hand, may be encouraged to regularly administer STAR Reading tests with a degree of fidelity and then carefully and thoughtfully analyze the data as one basis for meaningful instructional adjustments, knowing that there is a positive correlation between STAR Reading and SBA ELA/Literacy scores. These findings may also spur teachers to buy into the idea of closely examining the STAR Reading data in order to identify students who are at risk of not meeting literacy standards and in response to the data mediate these students' learning experiences so that they can become more proficient readers by the time they finish third grade.

Access to a short and efficient reading progress monitoring test like STAR Reading, which, as pointed out in this study, is statistically significantly correlated to the Washington State standardized exam measuring students' reading, writing, listening, speaking, and research skills, should greatly facilitate efforts to help students meet state literacy standards at the end of the year. Because the STAR Reading results are made available immediately after the test is completed, teachers and educational leaders can more readily make instantaneous instructional decisions, making the response to identified academic needs more timely, and therefore, more helpful, constructive, and efficacious. It is important to note, however, that progress monitoring measures should not displace the use of formative assessments in the classroom. Formative assessments play a critical role in student learning and serve a purpose that complements that of a progress monitoring measure. They furnish information about student learning that is necessary to make fluid, ongoing adjustments to a specific lesson in order to ensure that students are learning the skills and concepts on hand. Responsible use of data in order to make informed decisions about teaching and learning should utilize a comprehensive assessment system that consists of various assessment types; therefore, the use and analysis of progress monitoring data should complement, not supplant, formative assessment data.

Research Question 2. Is there a statistically significant relationship between the students' STAR Reading assessment scores from fall, winter, and spring and their SBA English Language Arts/Literacy scores?

The researcher hypothesized that there is a statistically significant relationship between the students' STAR Reading assessment scores from fall, winter, and spring and

their SBA ELA/Literacy scores. The results of the hierarchical multiple regression analysis confirmed this hypothesis. The basis for hierarchical multiple regression is to establish whether a set of independent variables can predict the outcome variable or explain the variance in the dependent variable. With this fundamental goal in mind, hierarchical regression results for this research question showed that the regression model containing the STAR Reading fall, winter, and spring scores explained 69.0% ($R^2 = .690$) of the variance in the students' SBA ELA/Literacy scores to a statistically significant level ($p < .001$). Taking this basic goal a step further, hierarchical multiple regression allowed the researcher to determine the order of the inclusion of the independent variables into the regression model in order to help ascertain the relative importance of a predictor based on hierarchical relevance. Results showed that the STAR Reading spring scores accounted for 62.3% ($R^2 = .623, p = .000$) of the variance in the SBA ELA/Literacy scores. Because the STAR Reading spring scores account for over 60% of the variance in students' SBA ELA/Literacy scores, it might be worthwhile to consider giving the STAR Reading test as far in advance of the SBA as possible so that teachers have ample time to act on the results and address whatever lingering reading gaps and deficiencies students may have prior to taking the SBA.

Adding the STAR Reading fall and winter scores to the hierarchical regression model, the value of R^2 increased from .623 to .690 ($\Delta R^2 = .067, p = .000$). This means that the inclusion of STAR Reading fall and winter scores into the model explained a small but statistically significant variation in SBA ELA/Literacy scores over and above the STAR Reading spring scores. The values of the standardized coefficient ($\beta = .34, \beta = .26$, and $\beta = .29$ for STAR Reading spring, winter, and fall scores respectively) also

suggest that the three test administration periods have comparable levels of importance when it comes to their contribution to the regression model. The STAR Reading spring score has the highest standardized coefficient of the three testing periods, which is not surprising, considering that it is administered late in the year after the students have received ample reading and literacy instruction. However, the standardized coefficient value for STAR Reading spring scores is not that drastically different from that of STAR Reading fall and winter scores. These findings suggest that teachers should not overlook the scores from the STAR Reading fall and winter test administrations and focus solely on the spring scores, because they do statistically significantly contribute to the variance in SBA ELA/Literacy scores. Needless to say, students are better served when their teachers depend upon a comprehensive assessment system to gather salient information about their academic progress.

Research Question 3. Is there a statistically significant relationship between the students' SBA English Language Arts/Literacy scores and their STAR Reading spring scores, gender, socioeconomic status, and SPED status?

For this research question, the researcher hypothesized that there is a statistically significant relationship between SBA ELA/Literacy scores and STAR Reading spring scores, gender, Free and Reduced Lunch status, and Special Education status. The multiple regression results confirmed this hypothesis ($R = .831$, $R^2 = .641$, $F(4, 552)$, adjusted $R^2 = .639$, $p = .000$).

The multiple regression results for this research question showed that of the four predictor variables (STAR Reading spring score, gender, Free and Reduced Lunch status, and Special Education status), the independent variable with the strongest relationship to

the SBA ELA/Literacy score was the STAR Reading spring score. This result is consistent with the results of the first two research questions. The first research question showed a strong, statistically significant relationship between the STAR Reading and SBA ELA/Literacy scores, while the second research question showed STAR Reading spring scores as contributing the most to the hierarchical regression model compared to STAR Reading fall and winter scores. These consistent findings shed light on the value, importance, and utility of STAR Reading spring scores as part of a comprehensive assessment plan used in curriculum and instruction. As stated previously, a short, practical, and efficient progress monitoring measure that is positively correlated with the SBA ELA/Literacy assessment should be very useful and beneficial to teachers in identifying at-risk readers and allowing them to provide targeted, differentiated, and responsive literacy instruction and intervention, especially when these results are analyzed along with other assessment results such as classroom-based formative assessments.

One finding from the multiple regression results that is not at all surprising, given the amount of available research and literature, has to do with the SBA ELA/Literacy performance of female students compared to their male classmates. Findings from the multiple regression analysis indicated that, all other factors being equal, female students' SBA ELA/Literacy scores were on average 22.42 points higher than that of male students in third grade. This is consistent with literature and research about the gender gap in reading achievement. The SBA performance of students in this study who qualified for Free and Reduced Lunch compared to those who did not is also consistent with literature and research related to the income achievement gap. Third grade students participating in

this study who qualified for Free and Reduced Lunch scored 11.87 points lower, on average, on the SBA ELA/Literacy assessment compared to their peers who came from non-low-income families, with all other factors held constant. Because of a small sample of students who qualified for Special Education, the regression results for this group of students were statistically insignificant. Based on these above mentioned results, one can conclude that male students and students who qualified for free and reduced lunch were at a disadvantage compared to their female counterparts and non-low-income peers. It behooves teachers then to continue to be vigilant in mediating learning experiences for male and low-income students to help them improve their reading proficiency. Furthermore, schools and school districts should strive to develop a research-based intervention plan that may help teachers address learning gaps and the needs of identified struggling readers so that they can allocate or reallocate financial and human resources towards intervention strategies as necessary. Educational leaders at the building and district level may also benefit from a careful analysis of the STAR Reading data in order to provide targeted and sustained professional development that enhances teachers' ability to meet the specific needs of struggling readers.

Connections to Previous Research

This particular research study is connected to previous research and literature germane to the reading achievement gap, the impact of reading on academic achievement, and the use of progress monitoring measures. These connections are discussed below.

Reading achievement gap. As noted in the introduction of this paper, a reading achievement gap exists among student ethnic groups, between male and female students,

between low-income and non-low-income students, and between students who qualify for Special Education services and those who do not. In the 2011 and 2015 NAEP reading tests, minority students and students from low-income families did not perform as well as their peers (National Assessment of Education Progress, 2011, 2015). Meanwhile, there remains a gap between boys' and girls' reading skills (Clark & Burke, 2012; Loveless; 2015; Walker, 2015). The results of this particular study are in line with trends across the country and internationally when it comes to the achievement gap in reading. Findings from this study's multiple regression analysis show that with all other factors being equal, third grade boys did not perform as well as third grade girls on the SBA ELA/Literacy assessment, while third graders from low-income families also scored lower on the SBA than their non-low-income peers. These findings add to existing research and literature concerning the achievement gap in reading and highlights the need for educators to continue to be intentional with their efforts to close the achievement gap. Educators need to have a systematic intervention plan in place for struggling readers, especially low-income and minority students who may not have access to educational resources or parental support at home that are important in addressing reading proficiencies. With the dire consequences of reading deficiencies impacting students later in their academic years and in life, educators have the moral and ethical obligation to make reading proficiency for all students an urgent goal and to do all they can to achieve it.

Special Education status and student ethnicity have also consistently been cited in research and literature as factors impacting student literacy and overall academic achievement. In this study, the students' Special Education status was included as a variable in the multiple regression model; however, it did not yield any statistically

significant results, most likely due to a small sample size. The researcher did not include student ethnicity in the regression model as one of the predictor variables because the participants in this study were primarily Caucasian students. There were too few students in each ethnic subset to constitute a large enough sample size to yield statistically significant findings.

Impact of reading on academic achievement. Included in this paper's literature review were literature and research studies that point to the impact of reading proficiency on academic achievement. Many reading and literacy experts maintain that reading has a very strong influence on student learning and academic success, whether it is in the area of writing, English, math, science, or students' overall grades (Lonigan & Phillips, 2015; Nation & Norbury, 2005; O'Connor & Klein, 2004). For instance, experts and empirical evidence suggest that reading is the foundation for effective writing and that it is nearly impossible to achieve proficiency and sophistication as a writer without first achieving proficiency as a reader (Fletcher & Portalupi, 1998; Loban, 1963; Olness, 2005; Stotksy, 1983). Vilenius-Tuohimaa et al. (2008) claimed that reading skills and mathematical word problem skills are statistically significantly correlated, while Cromley (2009) asserted that reading comprehension leads to higher achievement in science.

The results of this present study adds to existing literature and research about the impact of reading on academic achievement. Spearman rank-order correlation and regression analyses showed that the STAR Reading test, which measures students' growth and level of reading proficiency, is highly correlated with the SBA English Language Arts/Literacy test, which assesses not only reading but also writing, listening, speaking, and research skills. This correlation suggests that reading proficiency is linked

to other aspects of literacy outside of reading, such as writing, speaking, listening, and research skills. Students' reading abilities are associated with how well they write, communicate and listen, and are able to complete research-related tasks. These findings illustrate the significance and urgency of ensuring that all students achieve reading proficiency, lest they find themselves struggling not only in reading but in other academic and literacy areas as well. Teachers should be given relevant, ongoing, sustained, and job-embedded professional development opportunities that enable them to strengthen their practice around literacy, effectively use data to inform instruction, and incorporate reading into all content areas. There must be a concerted effort within buildings and districts to allow teachers to collaborate with, regularly meet, and observe each other teach for the purpose of learning about effective literacy instruction and the integration of reading into all content areas. Systematic reading intervention plans that can be implemented alongside classroom-level efforts must also be in place.

Reading progress monitoring measures and standardized exams. A number of research studies that explored the relationship between standardized reading tests and interim assessments such as progress monitoring measures, benchmark assessments, or curriculum-based measures have been conducted over the years (McGlinchey & Hixson, 2004; Miller et al., 2015; Shapiro et al., 2006; Weinstein, 2011; Wood, 2006). The results of this present study revealed a statistically significant relationship between a standardized exam (Smarter Balanced Assessment or SBA) and a progress monitoring measure (STAR Reading), adding to the existing body of research. This is notable because an interim assessment that has a statistically significant relationship with a standardized test may be useful to teachers in the identification of students at risk of not

meeting state learning standards and may facilitate efforts to adjust instruction. Progress monitoring measures like STAR Reading are short and efficient, providing teachers with timely information that they can readily use, along with formative assessment data. Results from interim assessments like the STAR Reading test may also be used as a basis for decisions regarding targeted intervention systems, allocation and redirection of financial and human resources, and professional development in order to address identified reading needs, gaps, and deficiencies.

Limitations of the Study

There are several limitations to this study, which include the following: length of data collection, missing data, limited school district participation, and use of non-parametric test.

Length of data collection. Development of early literacy begins in kindergarten and goes all the way to second and third grade. Therefore, it is safe to assume that whatever early literacy instruction the study's participants received from kindergarten through second grade may have had an impact on their third grade reading assessment scores. Because the reading data collected for this particular study was only for one school year, 2014-2015, during the participants' third grade year, this study does not take into consideration the influence of K-2 instruction on the participants' third grade literacy. This is an inherent issue because not accounting for prior instruction makes it challenging to truly ascertain how much of the study's predictor variables truly affect the SBA results and how much is actually due to their reading skills entering third grade. Collecting data for only one school year is also a limitation because it merely provides a snapshot rather than a trend. In order to more fully establish the feasibility and utility of a

progress monitoring measure in predicting an outcome variable like the SBA, it is helpful to conduct a longitudinal study and examine the impact of the progress monitoring measure on the SBA over a period of time. Because of this limitation, the results presented in this study should be interpreted with caution.

Missing data. The sample size for the study comprised of 651 third graders enrolled in a semirural, medium-sized school district in the Pacific Northwest in the school year 2014-2015. Unfortunately, there were missing values in the data set because some students did not take one or more of the tests explored in this study (STAR Reading administered in the fall, winter, and spring of the 2014-2015 school year and the SBA English Language Arts/Literacy administered in the spring of 2015). Up to 14 percent of the full data set was missing. Missing data in quantitative studies can have serious implications and may result in statistical concerns such as decreased statistical power, increased standard errors, weakened generalizability of findings, and biased results (Dong & Peng, 2013). The researcher handled the missing data by excluding any cases with missing values from the statistical analyses; this procedure is called listwise deletion. A major concern that researchers encounter when removing cases with missing data is reduced sample size, which may lead to biased estimates or statistically insignificant findings. However, for this particular study, the sample size remained large enough, which provided a level of assurance that the statistical analyses used in the study would not yield biased results. As a result, listwise deletion was used instead of using other methods that may have brought about more significant issues such as an overfitting of the statistical model or not preserving relationships between variables. Although listwise

deletion appears to be an acceptable method of handling missing data, results of this study need to be interpreted with caution.

Limited school district participation. Because it is difficult to obtain specific student data from school districts, the data collection in this proposed study was confined to one semirural, medium-sized school district in the state of Washington. This posed a limitation because the study did not have a sample that was fully representative of the state's student population. The state has differing school district compositions (rural, semirural, urban, suburban) and sizes (small, medium, large), and this study does not reflect that diversity. The study's sample is also not representative of the state's student population in terms of student heritage and income. The sample is much less diverse than the state's population and certainly does not have as many low-income families. More specifically, the state of Washington in 2014-2015 had approximately 57% of its student population made up of Caucasian students, while the participating school district was much less diverse with over 78% Caucasian students and less than 22% minority students. The state had about 45% of its entire student population come from low-income families in 2014-2015, but the participating school district only had 21% of its students qualify for free and reduced lunch. Therefore, even though the sample size in this study was large ($N = 651$), the sample mean may not be truly representative of the population mean. This makes the study's generalizability, or the application of the study's findings and conclusions from a sample population to the larger population, problematic.

Replication of this study that involves participation of the different types of schools districts in the state (rural, semirural, urban, suburban, small, medium, large) would be helpful in establishing research generalizability.

Recommendations for Further Study

The results of this study add to current research and literature around the relationship between a summative standardized exam and a progress monitoring measure, specifically demonstrating a statistically significant relationship between the SBA and STAR Reading. However, data collected for this study came from only one school year. Future longitudinal studies that include data taken over a period of time may be helpful in more fully establishing the relationship between the two and to ascertain if a pattern emerges and the relationship between the SBA and STAR Reading holds true over an extended period of time. A relationship between these two variables over several years may further substantiate the value and utility of STAR Reading scores as a variable that explains the variance in students' SBA scores.

Future research that takes into account students' reading proficiency and learning experiences prior to third grade is also recommended. The present study collected assessment data from a cohort of third graders from the 2014-2015 school year and did not address the students' reading proficiency and literacy instruction that they received in previous years that may have had an influence on their year-end summative test performance. Future studies that account for previous literacy experience and skill can help researchers better determine the extent of the unique contributions of the STAR Reading scores to the outcome variable and can provide researchers with a certain degree of assurance that that variability in the SBA scores were not influenced by students' reading proficiency coming into third grade.

Previous research and literature has contended that minority students lag behind their peers in reading, so ideally student ethnicity should have been included as one of the

predictor variables in this particular study. Unfortunately, this was not possible because of a lack of diversity in the present study's sample. Future research that includes student ethnicity as a predictor variable is recommended. Understanding how much of the variance in a student's SBA score is explained by their ethnicity allows teachers to better tailor their instruction to meet the specific reading needs of minority students. This type of information would be beneficial to teachers and educational leaders as they leverage culturally rich learning experiences, content, and reading materials to target a student's specific reading deficiency.

This study did not have any statistically significant findings related to Special Education students because of the small sample of that particular group of students participating in the study. Future research that includes SPED status as an independent variable is recommended. Knowing the variance in the SBA score that can be explained by a student's SPED status would provide helpful information that teachers can use to better develop individualized learning plans that lead to personalized, targeted, and effective modifications and accommodations to instruction.

To properly address issues pertaining to generalizability, future studies that include more school districts that are representative of the state's student population are warranted. The present study used convenience sampling and collected data from only one school district whose student population does not mirror the state's student population in terms of ethnic diversity and income. Therefore, generalizability of this particular study is problematic.

The current study simply addressed the relationship between STAR Reading and SBA English Language Arts/Literacy to see how much of the variance in the SBA score

can be explained by the progress monitoring measure. While this is certainly important information that can have significant implications on reading instruction, intervention, and enrichment, future studies that delve into the efficacy of instructional decisions and changes implemented by teachers and educational leaders in response to the data would be extremely insightful, informative, and instructive. Having access to interim assessment data that may predict student performance on the summative assessment is beneficial; however, what one does in response to the data is even more significant. This kind of study may shed light on the barriers and challenges that teachers face in intentionally and thoughtfully analyzing data and implementing data-informed changes to their instruction. It may also give teachers and educational leaders helpful information on what intervention works and what does not. With public resources being very scarce in the midst of an ever growing set of demands and needs, schools can ill afford to expend limited resources on instructional strategies, intervention systems, and practices that have no empirical basis to support their efficacy.

References

- Adams, M. J., Anderson, R. C., & Durkin, D. (1978). *Beginning reading: Theory and practice*. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/17475/ctrstreadeducrpepv01977i00003_opt.pdf?sequence=1
- Berry, W. D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.
- Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal*, 29(5), 623-625.
- Briggs, A., Clark, T., Wolstenholme, J., & Clark, P. (2003). Missing...presumed at random: Cost-Analysis of incomplete data. *Health Economics*, 12, 377-392.
- Brookhart, S. M. (2004). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, 100, 5-14.
- Center for Public Education. (2015). *Learning to read, reading to learn: Why third grade is a pivotal year for mastering literacy*. Retrieved from <http://www.centerforpubliceducation.org/earlyliteracy>
- Clark, C., & Burke, D. (2012). *Boys' Reading Commission: A review of existing research to underpin the Commission*. London, England: National Literacy Trust.
- Clay, M. M. (1985). *The early detection of reading difficulties* (3rd ed.). Auckland, NZ: Heinemann Education.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.
- Common Core State Standards Initiative. (2016). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and*

- Technical Subjects*. Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf
- Connecticut State Department, (n.d.). *Early literacy development: A focus on preschool*. Retrieved from <http://www.sde.ct.gov/sde/lib/sde/PDF/DEPS/Early/literacy.pdf>
- Cromley, J. G. (2009) Reading achievement and science proficiency: International comparison from the Programme on International Student Assessment. *Reading Psychology, 30*(2), 89-118.
- Danielson, C., Axtell, D., Bevan, P., Cleland, B., McKay, C., Phillips, E., & Wright, K. (2009). *Implementing the framework for teaching in enhancing professional practice*. Alexandria, VA: ASCD.
- Danielson Group. (2013). *The Danielson framework for teaching*. Retrieved from <https://www.danielsongroup.org/download/?download=448>.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 150-161.
- Dong, Y., & Peng, C. J. (2013). Principled missing data methods for researchers. *SpringerPlus, 2*, 222. <http://doi.org/10.1186/2193-1801-2-222>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ...Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 428-1446.
- Durrant, G. (2005). *Imputation methods for handling item-nonresponse in the Social Sciences: A methodological review*. ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute, University of

- Southampton, NCRM Methods Review Papers. Retrieved from <http://eprints.ncrm.ac.uk/86/1/MethodsReviewPaperNCRM-002.pdf>.
- Espin, C., & Deno, S. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial & Special Education, 14*(6), 47.
- Feister, L. (2013). *Early warning confirmed: A research update on third-grade reading*. Baltimore, MD: Annie E. Casey Foundation.
- Feuerstein, R. (1990). The theory of structural cognitive modifiability. In B. Presseisen (Ed.), *Learning and thinking styles: Classroom interaction*. Washington, DC: National Education Association.
- Feuerstein, R. (2008). Conductive education and structural cognitive modifiability. *RACE, 1*, 5-8.
- Feuerstein, R., Falik, L. H., & Feuerstein, R. (2003). The learning potential assessment device. In Feuerstein, R. *Feuerstein's theory & applied systems: A reader* (pp. 54-97). Jerusalem, Israel: International Center for the Enhancement of Learning Potential.
- Feuerstein, R., Klein, P. S., & Tannenbaum, A. J. (Ed.) (1999). *The mediated learning experience (MLE): Theoretical, psychosocial and learning implications*. London, England: Freund.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed). Thousand Oaks, CA: Sage.
- Fletcher, R., & Portalupi, J. (1998). *Craft lessons: Teaching writing K-8*. Portland, MN: Stenhouse.

- Gambrell, L. B., Morrow, L. M., Neuman, S. B., & Pressley, M. (Eds.). (1999). *Best practices in literacy instruction*. New York, NY: The Guilford Press.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.
- Good, R. H. Simmons, D. C., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators for foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.
- Henderson, S., Petrosino, A., Guckenberger, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007-No. 039). Washington DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from <https://ies.ed.gov/ncee/edlabs>
- Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. Retrieved from <http://www.aecf.org/m/resourcedoc/AECF-DoubleJeopardy-2012-Full.pdf>.
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30, 1-15.

- Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, *61*(1), 79–90. <http://doi.org/10.1198/000313007X172556>
- Howell, K. W., & Nolet, V. (1999). *Curriculum-based evaluation: Teaching and decision-making* (3rd ed.). Belmont, CA: Wadsworth.
- Juel, C. (1998). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437-447.
- Laerd Statistics. (2015). Multiple regression using SPSS Statistics. *Statistical tutorial and software guides*. Retrieved from <https://statistics.laerd.com>
- Lesnick, J., Goerge, R., Smithgall, C., & Gwyne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment?* Chicago, IL: Chapin Hall at the University of Chicago.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202.
- Loban, W. (1963). *The language of elementary school children* (Research Report 1). Urbana, IL: National Council of Teachers of English.
- Lonigan, C. J., & Phillips, B. M. (2015). Response to instruction in preschool: Results of two randomized studies with children at significant risk of reading difficulties. *Journal of Educational Psychology*, *108*(1), 114-129.
- Loveless, T. (2015). The 2015 Brown Center Report on American education: How well are American students learning? With sections on the gender gap in reading,

- effects of the Common Core, and student engagement. *Brown Center on Education Policy at Brookings*, 3(4).
- Mart, C. T. (2012). Developing speaking skills through reading. *International Journal of English Linguistics*, 2(6), 91-96.
- Marzano, R. J., Pickering, D., & McTighe, J. (2005). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, VA: ASCD.
- McDonald, J. H. (2014). *Handbook of Biological Statistics* (3rd ed.). Baltimore, MD: Sparky House.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33(2), 193-2003.
- Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, 77(1), 103-117.
- Miller, K. C., Bell, S. M., & McCallum, R. S. (2015). Using reading rate and comprehension CBM to predict high-stakes achievement. *Journal of Psychoeducational Assessment*, 33(8), 707-718.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M., G., Mallinckrodt, C., & Carroll, R., J., (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3), 445-464. <http://doi.org/10.1093/biostatistics/kxh001>
- Mooney, M. E. (1990). *Reading to, with, and by children*. Katonah, NY: Richard C. Owen.

Nation, K., & Norbury, C. F. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders, 25*, 21-32.

National Assessment of Educational Progress. (2011). *The Nation's Report Card*.

Retrieved from

<https://nces.ed.gov/nationsreportcard/subject/publications/stt2015/pdf/2016008WA4.pdf>

National Assessment of Educational Progress. (2015). *2015 Reading State Snapshot Report*. Retrieved

from <https://nces.ed.gov/nationsreportcard/subject/publications/stt2015/pdf/2016008WA4.pdf>

National Association for the Education of Young Children. (1998). Learning to read and write: Developmentally appropriate practices for young children. *Young Children, 53*(4), 30-46.

National Board for Professional Teaching Standards. (1989). *What teachers should know and be able to do: National Board for Professional Teaching Standards*.

Retrieved from

http://www.nbpts.org/sites/default/files/what_teachers_should_know.pdf

National Conference of State Legislatures. (n.d.). *Literacy and No Child Left Behind (NCLB)*. Retrieved from <http://www.ncsl.org/research/education/literacy-no-child-left-behind.aspx>

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Reading Panel*. Washington, DC: National Institute of Literacy.

- O'Brien, E. M. (2008). *From beginning to stellar: Five tips on developing skillful readers*. Alexandria, VA: Center for Public Education.
- O'Connor, I. M., & Klein, P. D. (2004). Exploration strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 34*(2), 115-127.
- Oller, J. (1979). *Language tests at school*. London, England: Longman.
- Olness, R. (2005). *Using literature to enhance writing instruction* (pp. 1-9). Newark, DE: International Reading Association.
- Pellegrino, J. (2003). Knowing what students know. *Issues in Science and Technology, 19*(2), 48-52.
- Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Retrieved from <http://www.achieve.org/files/TheRoleofInterimAssessments.pdf>
- Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development, 36*, 9-22.
- Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*(4), 353-383.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: ASCD.
- Raymond, P. (1988). Cloze procedure in the teaching of reading. *TESL Canada Journal, 6*(1), 91-97.
- Renaissance Learning. (n.d.). *Skills for reading learners*. Retrieved from <http://www.renaissance.com/Products/Star-Assessments/Star-Reading/Skills>

- Renaissance Learning. (2011). *Making RTI work: A practical guide to using data for a successful response to intervention program*. Retrieved from <http://doc.renlearn.com/KMNet/R004336320GJBD8A.pdf>.
- Renaissance Learning. (2012). *STAR Reading: Technical manual*. Retrieved from http://www.doe.virginia.gov/school_finance/procurement/student_growth_assessments/renaissance/STAR-Reading-Technical-Manual.pdf
- Renaissance Learning. (2014). *The research foundation for STAR assessments: The science of STAR*. Retrieved from <http://doc.renlearn.com/KMNet/R003957507GG2170.pdf>
- Renaissance Learning. (2015a). *Parent's guide to STAR Assessments: Questions and answers*. Retrieved from <http://doc.renlearn.com/KMNet/R0054872491706A8.pdf>
- Renaissance Learning. (2015b). *Relating STAR Reading and STAR Math to Washington Smarter Balanced Assessments performance*. Retrieved from <http://doc.renlearn.com/KMNet/R004581714GK7F0B.pdf>
- Reutzel, D. R. (2015). Early literacy research: Findings primary-grade teachers will want to know. *The Reading Teacher*, 69(1), 14-24.
- Roskos, K. A., Christie, J. F., & Richgels, D. J. (n.d.). *The essential of early literacy instruction*. Retrieved from <https://www.naeyc.org/files/yc/file/200303/Essentials.pdf>
- Runge, J. T., Lillenstein, D. J., & Kovaleski, J. F. (2016). Response to intervention and accountability systems. In S. R. Jimerson, M. K. Matthew, & A. M.

VanDerHeyden (Eds.). *Handbook of response to intervention: The science and practice of multitiered systems of support* (2nd ed.). New York, NY: Springer.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006).

Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, 24(1), 19-35.

Smarter Balanced Assessment Consortium. (2015). *Content specifications for the summative assessment of the Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and technical subjects*. Retrieved from https://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Content_Specs.pdf

Soley-Bori, M. (2013). *Dealing with missing data: Key assumptions and methods for applied analysis*. Boston University School of Public Health. Technical Report.

Retrieved

<http://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf>

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stotsky, S. (1983). Research on reading/writing relationships: A synthesis and suggested directions. *Language Arts*, 60(5), 627-642.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, 30, 73-87.

- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R.,...Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*, 468-479.
- Walker, T. (2015). Have boys caught up to girls in reading? Not yet, but the gap is narrowing. *NEA Today*. Retrieved from <http://neatoday.org/2015/04/02/have-boys-caught-up-to-girls-on-reading-not-yet-but-the-gap-is-narrowing/>
- Washington State Office of the Superintendent of Public Instruction. (2016). *State testing*. Retrieved from <http://www.k12.wa.us/assessment/StateTesting/FAQ.aspx>
- Vilenius-Tuohimaa, P., Aunola, K., & Nurmi, J. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28*(4), 409-426.
- Vogt, W. P. (2005). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage.
- Vygotsky, L. (1978). Interaction between learning and development. In Gauvain, M. & Cole, M. (Eds.), *Readings on the development of children* (2nd ed., pp. 29-36). New York, NY: W. H. Freeman and Company.
- Weinstein, E. E. (2011). *A study of the relationship between elementary school students' performance on progress monitoring measures of oral reading fluency and reading comprehension to the Pennsylvania system of school assessment* (Doctoral Dissertation). Philadelphia College of Osteopathic Medicine. Retrieved from http://digitalcommons.pcom.edu/psychology_dissertations. (Paper 204)

William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

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

Appendix A

STAR Reading Skills and Domains

Foundational Skills	Reading: Literature	Reading: Informational Text	Language
Phonics and Word Recognition Inflectional Endings/Affixes	Key Ideas and Details Character Setting Plot Theme Summary Inference and Evidence	Key Ideas and Details Prediction Main Idea and Details Inference and Evidence Sequence Compare and Contrast Cause and Effect Connections and Relationships Summary	Vocabulary Acquisition and Use Word Relationships Structural Analysis Context Clues Real-life Word Connections and Applications Vocabulary in Context Antonyms Multiple-Meaning Words Synonyms Word Reference Figures of Speech
Fluency Purpose of Reading/Reading with Comprehension	Craft and Structure Point of View Structure of Literary Text Word Meaning Connotation Author's Word Choice and Figurative Language Integration of Knowledge and Ideas Modes of Representation Analysis and Comparison Range of Reading and Level of Text Complexity Range of Reading Text Complexity	Craft and Structure Text Features Authors' Purpose and Perspective Word Meaning Organization Author's Word Choice and Figurative Language Connotation Integration of Knowledge and Ideas Modes of Representation Analysis and Comparison Range of Reading and Level of Text Complexity Range of Reading Text Complexity	

Appendix B

Grade 3 SBA English Language Arts/Literacy Targets

Literary Text	Informational Text
<p>1 Key Details: Given an inference or conclusion, use explicit details and implicit information from the text to support the inference or conclusion provided.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.</p>	<p>4 Key Details: Given an inference or conclusion, use explicit details and implicit information from the text to support the inference or conclusion provided.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers. RL-17 Use information gained from illustrations (e.g. maps and photographs) and the words in a text to demonstrate understanding of the text (e.g. where, when, why, and how key events occur).</p>
<p>2 Central Ideas: Identify or determine a central message, lesson or moral and explain how it is conveyed in the text through key details, key events, or the sequence of events.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers. RL-2 Recount stories, including fables, folktales, and myths from diverse cultures; determine the central message, lesson, or moral and explain how it is conveyed through key details in the text.</p>	<p>5 Central Ideas: Identify or determine a main idea and the key details that support it.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers. RL-2 Determine the main idea of a text; recount the key details and explain how they support the main idea.</p>
<p>3 Word Meanings: Determine intended meanings of words, including words with multiple meanings, based on context, word relationships, word structure, with primary focus on determining the meaning based on context and the academic vocabulary common to complex texts in all disciplines.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.</p>	<p>6 Word Meanings: Determine intended meanings of words, including words with multiple meanings, based on context, word relationships, word structure, with primary focus on determining the meaning based on context and the academic vocabulary common to complex texts in all disciplines.</p> <p>Standards: RL-1 Ask and answer questions to demonstrate understanding of a text, referring explicitly to the text as the basis for the answers.</p>

Appendix C

Hierarchical Regression – Linear Relationship

The following scatterplot shows a linear relationship between dependent variable and the independent variables taken collectively. The horizontal band indicates linear relationship.

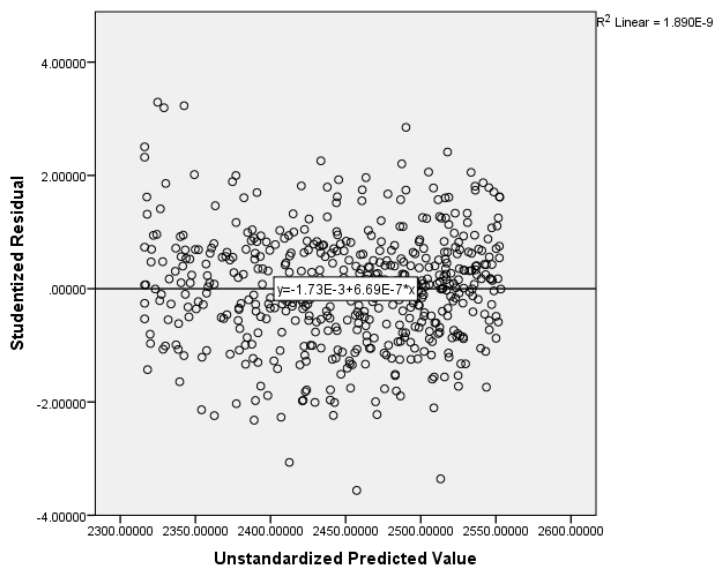


Figure C1. Relationship between SBA and STAR Reading

The following scatterplots show a linear relationship between the dependent variable and the different independent variables.

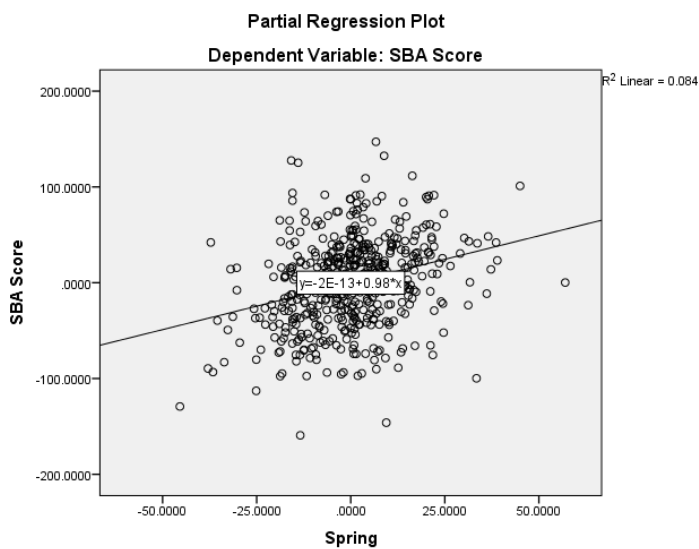


Figure C2. Relationship between SBA and STAR Reading Spring

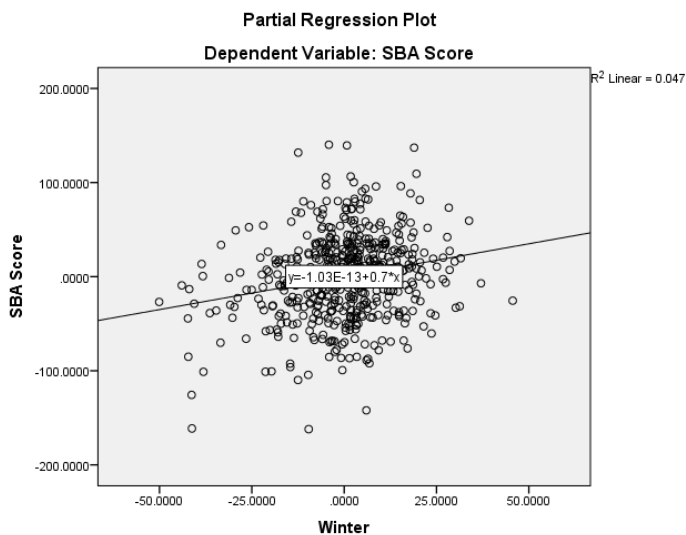


Figure C3. Relationship between SBA and STAR Reading Winter

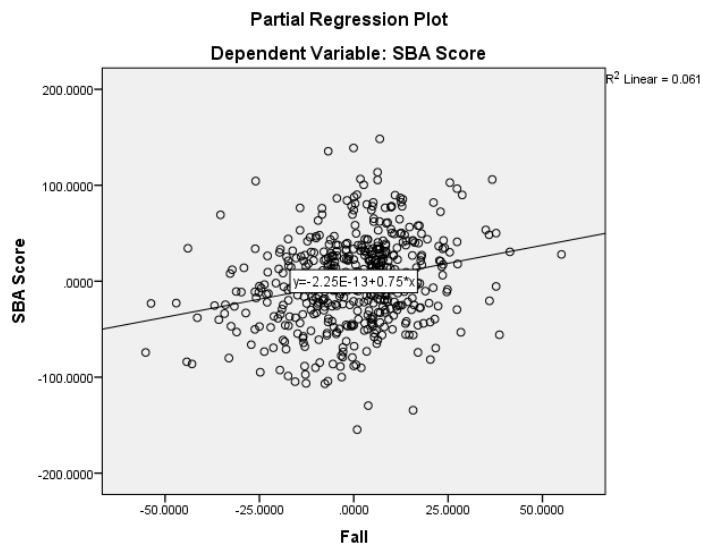


Figure C4. Relationship between SBA and STAR Reading Fall

Appendix D

Hierarchical Regression – Normality of Distribution

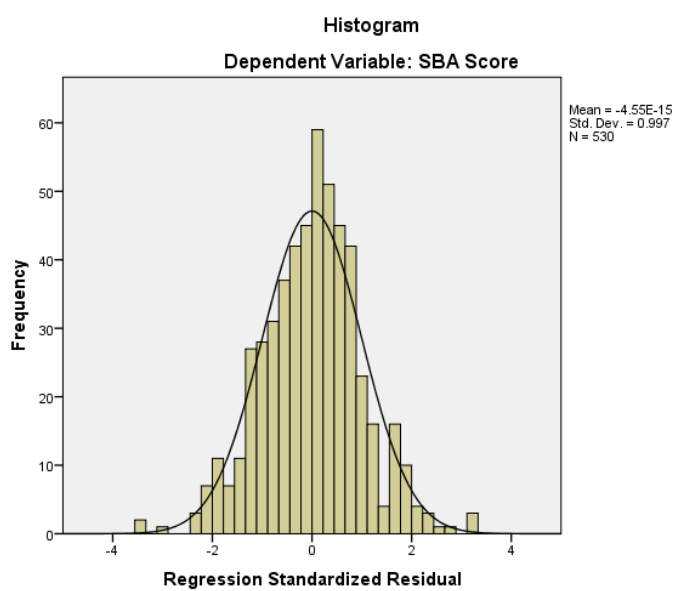


Figure D1. Histogram of residuals

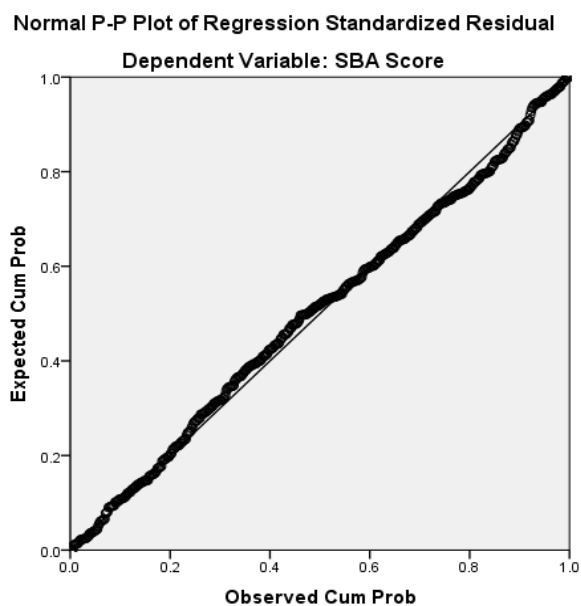


Figure D2. P-P Plot below of residuals

Appendix E

Multiple Regression – Linear Relationship

The following scatterplot shows a linear relationship between dependent variable and the independent variables taken collectively. The horizontal band indicates linear relationship.

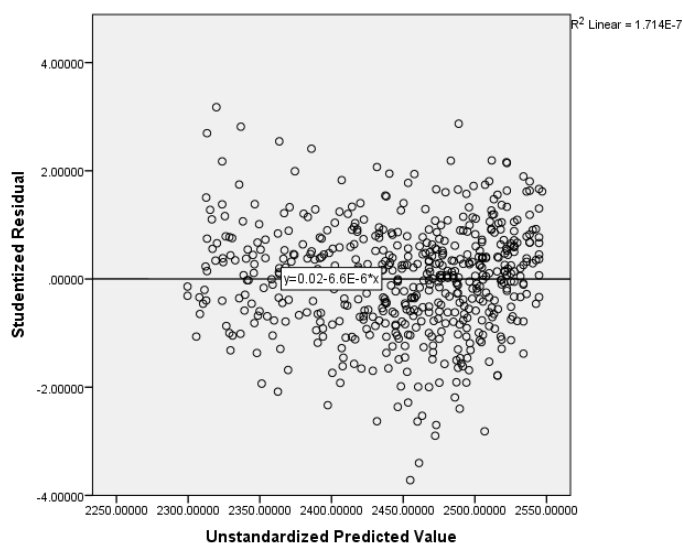


Figure E1. Relationship between SBA and STAR Reading Spring, Gender, SPED Status, FRL Status

The following scatterplots show a linear relationship between the dependent variable and the different independent variables.

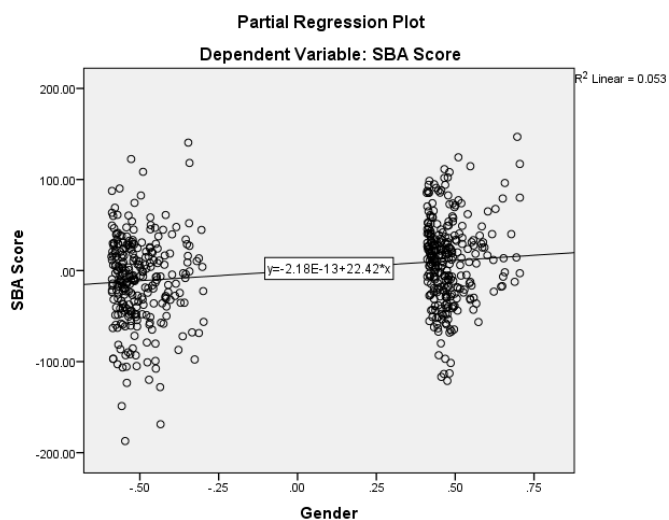


Figure E2. Relationship between SBA and gender

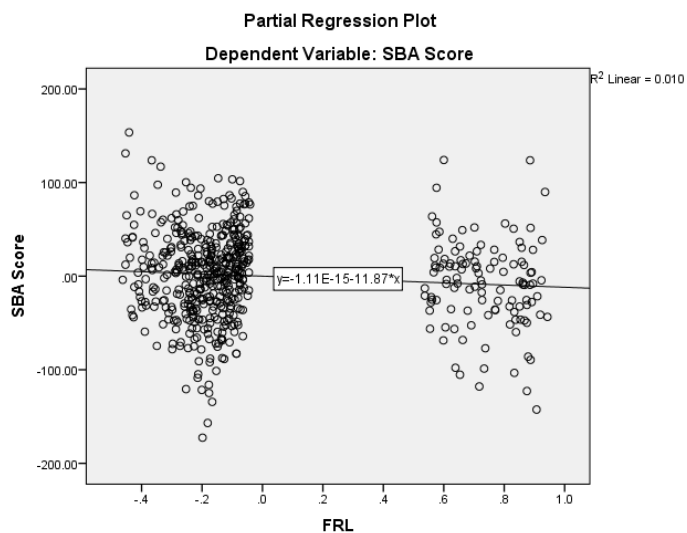


Figure E3. Relationship between SBA and FRL status

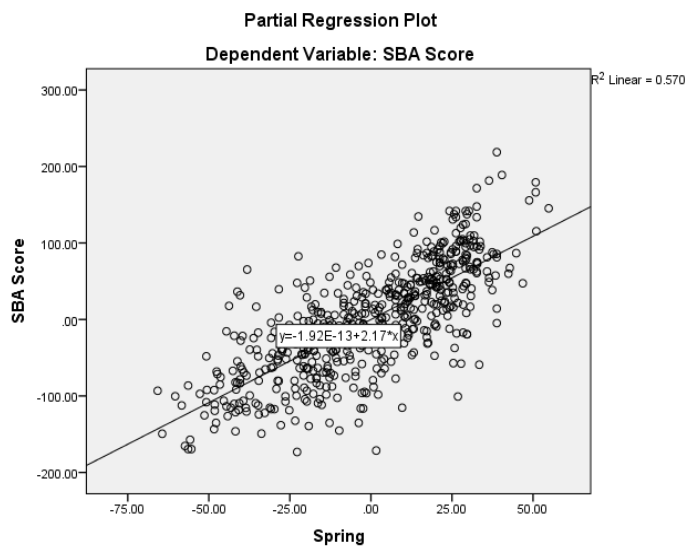


Figure E4. Relationship between SBA and STAR Reading Spring

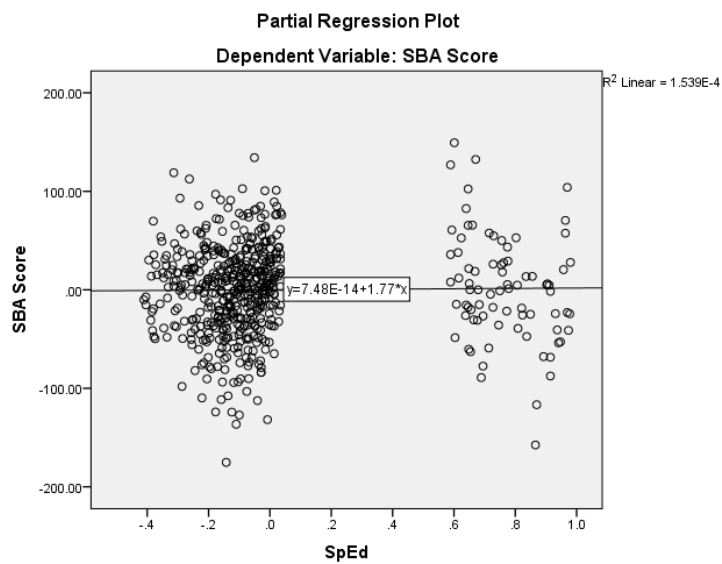


Figure E5. Relationship between SBA and SPED status

Appendix F

Multiple Regression – Normality of Distribution

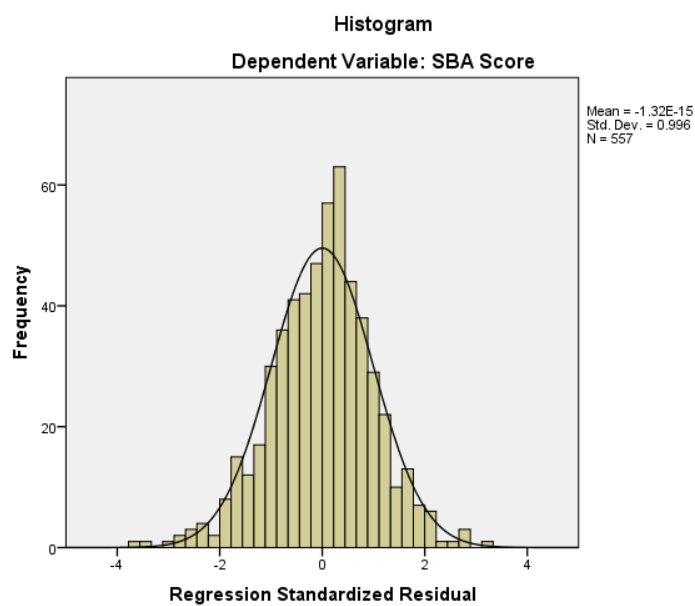


Figure F1. Histogram of residuals

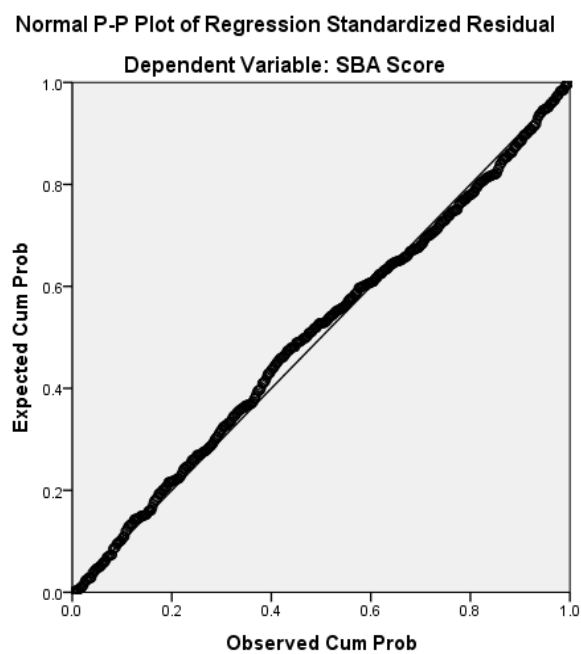


Figure F2. P-P plot of residuals

Appendix G
Institution Review Board

Response to: Miriam M. Mickelson
 Xc: Dr. Tom Alsbury
 Re: Exempt Review
 Subject: IRB Approval – IRB # 151603004 (Exempt)
 Date: March 29, 2016

Your research project **“The Relationship between Elementary School Students’ Performance on a Reading Progress Monitoring Measure and the Washington State Standardized Test”** has been approved under exempt IRB review. This study was approved under exempt review as it met the following criteria:

3. Research uses survey or interview procedures or observations (including observations by participants) of public behavior AND at least one of the following conditions exists:
 - a. Human participants cannot be identified directly or through identifiers code or numbers, **OR**
 - b. The participants’ responses or the observations recorded, if they became known outside research, cannot reasonably place the participant at risk of criminal or civil liability or be damaging to the participant’s financial standing or employment, **OR**
 - c. The research does not deal with sensitive aspects of the participant’s own behavior, such as illegal conduct, drug use, sexual behavior, or use of alcohol.

Your approval is in effect until 03/29/2017. Your study has been assigned IRB number: IRB # 151603004.

To complete your documents, please add your IRB # and expiration date to your study as needed.

If you need more than one year to complete data collection, you must file a request for an extension with me eight weeks before the expiration date of this study. Your request for an extension can be written or communicated through e-mail and must include a report on the status of your study. Otherwise you will need to file a new IRB application to continue with data collection after the expiration date.

Use your IRB number in any further communication regarding this study.

This is the only documentation that you will receive regarding your study’s approval. Please include it in your study’s documentation.

Best wishes on the completion of your research.

John L. Glancy

Member, SPU Institutional Review Board (IRB)