

Utah State University

DigitalCommons@USU

---

All Graduate Theses and Dissertations

Graduate Studies

---

5-2019

## Standards-Based Grading: A Correlational Study Between Grades and End-of-Level Test Scores

Tyler R. Poll  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Teacher Education and Professional Development Commons](#)

---

### Recommended Citation

Poll, Tyler R., "Standards-Based Grading: A Correlational Study Between Grades and End-of-Level Test Scores" (2019). *All Graduate Theses and Dissertations*. 7435.

<https://digitalcommons.usu.edu/etd/7435>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



STANDARDS-BASED GRADING: A CORRELATIONAL STUDY BETWEEN  
GRADES AND END-OF-LEVEL TEST SCORES

by

Tyler R. Poll

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Educational Leadership

Approved:

---

Courtney D. Stewart, Ph.D.  
Major Professor

---

Max L. Longhurst, Ph.D.  
Committee Member

---

Susan A. Turner, Ph.D.  
Committee Member

---

Michael Freeman, Ph.D.  
Committee Member

---

Robert W. Wagner, Ph.D.  
Committee Member

---

Laurens H. Smith, Ph.D.  
Interim Vice President for Research and  
Interim Dean of the School of Graduate  
Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2019

Copyright © Tyler R. Poll 2019

All Rights Reserved

## ABSTRACT

Standards-Based Grading: A Correlational Study Between Grades and End-of-Level  
Test Scores

by

Tyler R. Poll, Doctor of Philosophy

Utah State University, 2019

Major Professor: Courtney D. Stewart, Ph.D.  
Department: Teacher Education and Leadership

The purpose of this research study was to investigate the relationship between standards-based grades and end-of-level SAGE test scores. By using end of term grades, the researcher evaluated the correlation between grades and end-of-level SAGE test scores, the impact on student learning through the student growth percentile, and the differences in grades and test score data dependent on curriculum content area and years of teaching experience. Grade and SAGE data from the 2016-17 school year were obtained from the school district assessment department. The researcher analyzed the data using frequency histograms, multiple independent samples  $t$  tests, bivariate scatter plots, and linear regression models. In general, the independent samples  $t$  tests showed standards-based grades and traditional grades are both moderately correlated with student performance on the end-of-level SAGE tests. However, an examination of specific areas showed students who participated in standards-based classrooms earned higher grades,

scored higher on the end-of-level SAGE tests, and showed higher learning growth when compared to their peers who participated in traditional graded classrooms. Overall, the results suggested standards-based grading as a positive option for educators to consider for their grading practices. Therefore, educational leaders should consider providing standards-based grading professional development for their teachers to help benefit student learning and promote accurate grading practices.

(143 pages)

## PUBLIC ABSTRACT

Standards-Based Grading: A Correlational Study Between Grades and End-of-Level  
Test Scores

Tyler R. Poll

As students move from grade level to grade level and onto college, their grades have an impact on the number of opportunities available to students. The competition for entering college and earning a scholarship are at an all-time high and the grades students earn have a direct impact on future opportunities. Grading practices vary by teacher causing students' grades to mean different things.

Standards-based grading practices focus on removing teacher bias and puts emphasis on the learning students can demonstrate. Students are given assessments to determine learning and are given multiple opportunities to show what they have learned. Emphasis is placed a student's most current knowledge rather than an average of scores during the grading period.

This study focused on how student learning was impacted when secondary math, science, and language arts teachers use standards-based grading practices in their classrooms. Student learning was measured by term grades and end-of-level SAGE test scores. Results show students who attended a classroom with standards-based grades earned higher GPAs, performed better on the end-of-level test, and had more learning growth over the course of the school year, than their peers who participated in traditional grading classrooms.

## ACKNOWLEDGMENTS

The opportunity to further my education and participate in this learning process would not be possible without the many positive influences I have had in my life. First, I would like to thank Dr. Joe Matthews for encouraging me to pursue my doctoral degree. Although he retired early in my doctoral program, he left me in the excellent hands of Dr. Courtney Stewart. Dr. Stewart is an amazing educator, advocate, and friend. I hold him in the highest esteem possible. I would like to thank my teachers for dedicating their lives to teaching and Drs. Susan Turner, Robert Wagner, Max Longhurst, and Michael Freeman for serving on my doctoral committee. They have each influenced me positively in different ways.

I express appreciation to Dr. Logan Toone and Bryce Barth for their help with data. I am grateful for my classmates, especially Jennifer Throndsen, for the positivity and motivation. A big thanks to my colleagues and district leaders for the support and interest in my project. Thank you to my parents for teaching me the importance of education and for their support. I am grateful to my children who have sacrificed more than I ever would want them to. Most of all, I am grateful for my wife Lindsey, who has been supportive since day one and has carried more than her share of the load.

Tyler R. Poll

## CONTENTS

	Page
ABSTRACT.....	iii
PUBLIC ABSTRACT .....	v
ACKNOWLEDGMENTS .....	vi
LIST OF TABLES .....	ix
LIST OF FIGURES .....	x
CHAPTER	
I. INTRODUCTION.....	1
Statement of the Problem .....	7
Purpose .....	8
Research Questions .....	9
Assumptions .....	10
Scope and Limitations .....	10
Definition of Terms .....	10
Significance of the Study .....	11
Conclusion.....	12
II. REVIEW OF THE LITERATURE.....	13
Student Learning .....	14
1965 Elementary and Secondary Education Act .....	15
A Nation at Risk .....	16
No Child Left Behind Act .....	16
Common Core .....	17
Race to the Top.....	18
Assessment Overview .....	19
Grading.....	24
Standards-Based Grading.....	34
Grades Versus High-Stakes Testing Scores .....	46
Conclusion.....	47
III. METHODS.....	48
Research Questions .....	49



Research Design .....	50
Setting.....	50
Participants .....	50
Data Sources .....	53
Data Collection Procedures .....	53
Data Analysis Procedures.....	56
Conclusion.....	59
 IV. RESULTS.....	 60
Correlation Between Grades and End-of-Level SAGE Test Scores .....	61
Effect on Student Learning.....	66
Difference by Core Subject .....	67
Difference by Years of Teaching Experience .....	69
Conclusion.....	71
 V. DISCUSSION .....	 73
Overview .....	77
Review of the Research and Research Questions .....	78
Fidelity of Implementation.....	79
Discussion of Results .....	81
Practical Implications of the Study .....	94
Theoretical Implications of the Study .....	95
Limitations of the Research.....	96
Recommendations for Future Research .....	98
Conclusion.....	99
 REFERENCES .....	 104
 APPENDICES .....	 112
Appendix A: Standards-Based Grading in the Classroom .....	113
Appendix B: GPA and SAGE Error Bar Charts.....	118
Appendix C: SAGE $z$ Score Regression Residual.....	120
Appendix D: Error Bar Chart Showing Confidency Level for Residuals .....	122
Appendix E: Bar Graph Showing SGP Score Analysis .....	124
Appendix F: Error Bar Charts by Content .....	126
Appendix G: Error Bar Charts by Experience.....	129
 CURRICULUM VITAE.....	 132

## LIST OF TABLES

Table	Page
1. Common Differences Between Traditional Grading and Standards-Based Grading .....	34
2. Standards-Based Grading Cohort and Traditional Grading Group Demographics .....	51
3. Grade Point Average (GPA) Scale .....	57
4. Standards-Based and Traditional Grading Groups Descriptive Statistics for GPA and SAGE $z$ Scores .....	61
5. Regression Model: Descriptive Statistics for GPA and SAGE $z$ Scores .....	63
6. Regression ANOVA Model: GPA and SAGE $z$ Scores .....	63
7. Descriptive Statistics for GPA and SAGE $z$ Scores Based on Level of Implementation (Self-Identified) .....	65
8. Descriptive Statistics of Student Growth Percentiles for Both Group.....	66
9. Descriptive Statistics for SGP Based on Level of Implementation (Self-Identified).....	67
10. Content Area Statistics Regression Residuals-SAGE $z$ Score with GPA.....	68
11. Teaching Experience Group Statistics Regression Residuals-SAGE $z$ Score with GPA.....	70

## LIST OF FIGURES

Figure	Page
1. Standards-based grading mindset .....	81
B1. GPA analysis .....	119
B2. SAGE $z$ score analysis.....	119
C1. SAGE $z$ score regression residual .....	121
D1. Confidency Level for Residuals .....	123
E1. SGP sore analysis .....	125
F1. Language arts error bar chart: Confidency level for residuals .....	127
F2. Math error bar chart: Confidency level for residuals .....	127
F3. Science error bar chart: Confidency level for residuals .....	128
G1. 0-5 years teaching experience error bar chart: Confidency level for residuals.....	130
G2. 6-10 years teaching experience error bar chart: Confidency level for residuals.....	130
G3. 11-15 years teaching experience error bar chart: Confidency level for residuals.....	131
G4. 16+ years teaching experience error bar chart: Confidency level for residuals.....	131

## **CHAPTER I**

### **INTRODUCTION**

The tardy bell rings, and a smattering of students lazily walk toward their class. Teachers are beginning to take attendance as I, along with two other school administrators, clear the hallways and ensure each student arrives in their appropriate classroom. After scouring the halls for 5 minutes, I walk back to my office to respond to emails. Before I can sit down, the secretary enters and informs me a parent is waiting in the office and wants to talk to the principal immediately.

Curious as to the nature of this arrival, I respond, “Do you know who their child is and what they are here for?” Parents come to school to speak with a principal for a gamut of reasons and having an idea of what they want a few minutes before meeting with them can save a great deal of time.

The secretary gives me the student’s name and comments, “I don’t know what they want for sure, but they seem mad.” Despite this vagueness, I let her know I am ready to see the parent. As the secretary walks away, I quickly load our student information system on the computer and begin looking at the student’s records. I scour attendance, grades, and behavior, hoping to locate a clue as to why the parent is here and upset.

The student’s attendance looks sufficient, other than a few days missed due to illness and a trip to Disneyland. The student does not have any major disciplinary issues. He was reported for bullying a few weeks ago, but he and his parents denied any knowledge of him doing anything wrong. As I observe his grades, he has mostly A’s, a B+, and one F in math. Why does it always have to be math? I assume the failing grade in

math is the reason the parent is here and angry.

The culmination of this research occurs in the 45 seconds it takes the secretary to leave my office, retrieve the parent, and bring the parent back to my office to meet. As the parent enters, I introduce myself as the assistant principal and indicate I am pleased to meet her. It becomes clear my pleasantries are not welcome and she wants to get straight to business.

The mother begins by telling me her son has always been a straight-A student. Although he competes on multiple athletic competition teams and is extremely busy, he always turns in his homework. She lays the groundwork to let me know her son is nearly perfect and whatever her complaint is, it cannot possibly be her son's fault.

I smile politely and ask how I can help. She continues to explain how all his teachers love him and how well behaved he is—in and out of school. She explains that if he does not get an athletic scholarship, he will most definitely get an academic scholarship. The way she describes her son's achievement encourages me to believe he will receive both. She continues expressing how proud she is of her son. I can relate to that. I begin thinking about my own children and how much I love them. I cease thinking about my own children as she questions, "...and what are you going to do about it?"

I bring my focus back to reality and realize I had become distracted during her monologue. I ask a follow-up question to ensure I know what her concern is regarding. Thankfully, during my quick, 45-second research, I am confident she is here to discuss his math grade.

I pull up his math scores and begin to investigate why he is currently failing math.

His math scores for the most part look okay. I see the teacher has marked some assignments “late,” which I assume means he was docked credit for handing the assignments in after the deadline. I also see he is missing a few homework assignments; however, he has done reasonably well on the tests. I also observe he was given some zeros on a few quizzes. Being familiar with the teacher, I know she has a strict policy that quizzes cannot be made up if the student is absent. Attending class and participating is very important to her. The missing work is titled, “Homework 3/27, Homework 4/5, Quiz #4” and so on.

The mother inquires why her son is failing math and quite honestly, I do not know. I want to be able to give the mother an answer that would indicate he is failing because he has not met a standard or he is not proficient in a certain area. I want to be able to tell her that it is okay because the teacher is working with him until he learns the concepts. But I am unable to provide her with these reassurances. All I can do is look at the mother and say, “Your son is failing math because of some missing assignments. He also missed some quizzes while you and your family were out of town. According to the math teacher, those quizzes cannot be made up unless he has a doctor’s note.”

I know this is not a sufficient response to her question, but to be honest, I do not have the information to let this mother know what can be done for her son to earn a grade that demonstrates learning. The titles of the assignments, quizzes, and tests are labeled by date and in sequential order. They are not labeled by the math standard or what the child is actually learning. The mother looks at me in disgust and responds, “Well, how can he raise his grade?” I open my mouth to ask a question to which I already know the answer.

I have asked this question many times and usually get the same answer.

I ask, “What did the teacher say when you talked to her?” The mother tells me she has not talked to the teacher because the teacher “does not care” and “will not budge.” The mother wants me to march into the classroom and let the teacher know this child is an angel and he deserves an A. Essentially, she wants me to tell the teacher he needs an A because he is a good student, not because he has demonstrated proficiency in math skills and standards.

As the assistant principal, I do not have the answers, and I do not have the power to change a math grade. I therefore inform the mother we will need to schedule a meeting with the teacher, myself, parents, and the child to discuss this matter further. The purpose of the meeting will be to find a way for her son to demonstrate learning in the classroom and earn a better grade.

At the meeting a few days later, both parents are in attendance. Their sole purpose is to ensure their son has an A in math at the end of the term. The teacher is there to defend herself and her grading practices. I am there as a mediator to make sure everybody leaves in agreement. The teacher focuses on the student’s lack of attending class and the one time he was disruptive in class. The parents do not believe he was ever disruptive and they defend their trip to Disneyland as much needed family time. They only want to know what their son can do to raise his grade. Eventually the teacher budes and provides some extra credit opportunities, many of which will have nothing to do with the math skills and standards actually studied while he was on vacation.

During the discussion, it is clear to me what is wrong with this situation. The

focus of the meeting is on a letter grade, not what the student has learned. It is not the parents' fault; they went through the same school system. They know how the system works and they know how to get what they want. The fault is mine, the teacher's, and the school's. We are still relying on a system of assessing and grading that is outdated and ineffective. Something must change.

As the frequency of assessments increases and the results of high-stakes testing affects students and schools, it is important school leaders identify methods to help ensure students are being assessed appropriately. If assessments are provided effectively and feedback is given to students, more learning can take place (Phelps, 2012). The way teachers assess students has not changed for more than a century. Educators and educational leaders must look for new and innovative ways to assess student learning at their schools and help students prepare for college and their future career.

The duties of school leaders continue to grow as education adapts and reforms. While the responsibilities may seem overwhelming, ensuring student success is of the utmost importance. With standardized testing taking a bigger role in education, school leaders need to look at other measures of student learning. Administrative support is needed to help teachers develop effective classroom grading practices that are compatible with newer assessment methods and communicate the changes with educational stakeholders (Suurtamm, 2004). Aligning assessments and having continuity in grading and testing practices helps ensure valid and reliable data. Keeping grading and testing data consistent allow teachers and school leaders to identify students who did not meet the standards and allows educators to pinpoint what interventions a student needs to be



successful.

With the recent focus on common standards through Common Core, some educators have begun to look at changing how to assign grades and assess learning. For the majority, however, traditional grading continues to be a mainstay in educational practice. Traditionally, grades have been used to sort students into higher and lower achieving classes. Grades have been used to determine honor rolls and right of membership in honor societies, student government, and athletic teams (Guskey, 2015). Having high grades can provide many opportunities; having low grades limits students from pursuing the same opportunities. Grades can have an immense impact on a student's life. The education system believes that grades should reflect a student's growth and progress, yet it continues to use an outdated grading system that does not reflect the ability to effectively determine student learning.

Standards-based grading enables educators to evaluate student learning based on how well the student masters specific content and skills. Instead of averaging scores for a set amount of time, students are graded based on how they demonstrate knowledge in each respective standard (Hooper & Cowell, 2014). Clearly describing the expected standards of performance is important, that way students know what they are expected to learn (Buckles, Schug, & Watts, 2001). Using standards-based grading methods provides the opportunity for educators, students, and parents to know the specific standards and identify where students need more support and learning interventions. By using data from secondary teachers who are currently implementing standards-based grading procedures in their classroom and the correlation of grades with end-of-level test scores, this study

will provide educational leaders with data in an area that lacks research.

### **Statement of the Problem**

The traditional way of determining student learning and assigning grades is unreliable and commonly based off of each individual teacher's preferences. Determining what a student has learned and assigning a grade to that student can be complex. The grades students receive can literally mean the difference between graduating from high school and not graduating, attending college and not attending college. Because grades have an impact on students' futures, it would be logical to assume educators are trained in assessment and assigning grades. However, undergraduate and graduate courses rarely teach curriculum on effective student grading. As a result, educators seldom know about different grading practices, the pros and cons of each, or the effect grading practices have on students (Brookhart, 2011; Stiggins & Chappuis, 2011). Because of the lack of training and knowledge regarding grading practices, teachers most often grade their students using the same grading and reporting methods they remember their teachers using in elementary and secondary schools (Frary, Cross, & Weber, 1993; Guskey & Bailey, 2001).

The way grades are determined, valued, and even collected has been around for a long time, yet research does not show the traditional way of assigning grades to be an accurate reflection of student learning. In fact, as early as 1912, Starch and Elliot indicated 147 English teachers assigned widely varied grades to two identical papers from students. They found the focus of what should count on the grade varied based on

teacher expectations. Educators at the time argued good writing is subjective, so Starch and Elliot (1913) did a similar study with math assignments. One geometry paper was scored by 128 teachers and scores ranged from 28% to 95%. As with their previous study, Starch and Elliot found teachers gave points dependent of the learning indicators they were assigned to look for on the assessment.

Recent research has shown evidence used in grading practices varies considerably from teacher to teacher and is even inconsistent within a teacher's own practice (Brookhart, 1993; Howley, Kusimo, & Parrot, 2000). Duncan and Noonan (2007) found secondary teachers' grading practices varied by subject area and non-achievement factors were regularly calculated into students' grades. In 2011, Brimi replicated Starch and Elliot's 1912 study and found similar results. Brimi had 73 high school teachers, who had received twenty hours of training in a writing program, grade a paper on a 100-point grading scale. Scores ranged from 50% to 96%. The results were nearly identical to the study completed a century earlier. The traditional way of determining student learning through the use of grades is outdated, inconsistent, and unreliable (Brookhart, 2004; Marzano, 2000).

### **Purpose**

The purpose of this study is to evaluate the effectiveness of determining student learning through the use of standards-based grading in secondary classrooms and the correlation those grades have with end-of-level test scores. While grading practices vary across the world, most educators use the same grading practices that have been used for

over 100 years. With the recent emphasis on content standards, educators are looking for more effective ways to determine student learning. Little research has been done to determine if standards-based grading is an effective way to assess student learning. By looking at teachers from a large, suburban school district in the western U.S. who are beginning to implement standards-based grading in the classroom and comparing student grades with student test scores, results will add to the existing research for standards-based grading.

### **Research Questions**

There are four areas of focus in this study: (1) determine if standards-based grading is an effective way to assess student learning in comparison with end-of-level test scores, (2) determine if the use of standards-based grading increases student learning as measured by the standard growth percentage on end-of-level tests, (3) understand what relationships exist between math, language arts, and science classes implementing standards-based grading, and (4) determine if teachers' years of experience affects the correlation between grades and end-of-level test scores. In order to address the areas of focus identified by this study, the following four research questions guided this study.

1. Is there a positive correlation between standards-based grading term grades, and end-of-level SAGE test scores?
2. To what extent does the standards-based grading environment increase or decrease student learning in comparison with traditional grading, as measured by the student growth percentage on end-of-level SAGE test scores?
3. Is there a difference in the correlation of standards-based grades and end-of-level SAGE test scores between math, language arts, and science?
4. Does the correlation between standards-based grades and end-of-level test

scores vary based on teachers' years of teaching experience?

### **Assumptions**

The major assumption in this proposed study is the cohort of teachers implementing standards-based grading procedures represents the larger population. Another assumption is that survey respondents will answer the survey questions truthfully. Last, the researcher is assuming students will do their best on the end-of-level SAGE tests.

### **Scope and Limitations**

The proposed research is limited by a few factors. The first limitation is time. Ideally, it would be beneficial to look at data for several years to compile more reliable results. Although the sample size is large, having only one year of data is a limitation. Second, the study is limited by the researcher's knowledge of the teachers in the traditional grading group. Some of these teachers could be in the process of attempting to use different grading techniques instead of the traditional grading practices often seen in schools. This is a limitation because it combines a segment of the population with a group into which they may not fit.

### **Definition of Terms**

The following terms are used throughout this study.

*Reliability:* A measure of the degree to which a test is free from measurement

error (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999).

*Student Assessment for Growth and Excellence (SAGE)*. The Utah end-of-level computer adaptive assessment aligned to the State's language arts, mathematics, and science standards.

*Student Growth Percentile (SGP)*. Measures the amount of academic growth a student has made, as a result of one year of instruction, compared to similarly performing peers across the State. Average SGP is 50%

*Standards*. Learning goals for what students should know and be able to do at each grade level.

*Traditional Grading Group*. Set of math, science, and language arts teachers who did not participate in the standards-based grading cohort.

*Validity*. A measure of the degree to which a survey has evidence that supports the inferences made from the scores (AERA et al., 1999).

### **Significance of the Study**

Educators are frequently looking for ways to determine if students have learned concepts and standards taught within the classroom. Term grades and test scores are the indicators used most often to reflect student learning. The current method of grading used by most educators has been shown to be an unreliable way to determine if a student has learned and in which standards a student is deficient (Brimhi, 2011; Guskey, 2015; O'Connor & Wormeli, 2011; Starch & Elliot, 1913). The potential evidence from this

study would provide research showing standards-based grades are more accurately correlated with end-of-level tests and are more conducive to support student learning. The research on standards-based grading will continue to grow and will provide educators with data to inform their decisions about standards-based grading and its value in the classroom.

### **Conclusion**

Educators have been having the same conversation about student grades with parents and students for far too long. Educational stakeholders know how important grades are, but the letter grade has become the focus of education, rather than what a student has learned and can do. Focus must be put back on the skills and knowledge students have learned and can demonstrate. If research on grading continues to examine traditional grading methods rather than newer grading methods, grading in education will continue along the same cycle we have seen for the past 100 years. Developing meaningful and equitable grading practices is a challenge for educators at all levels. The challenge will be more overwhelming if educators continue to use the same methods. “We’ve always done it that way is insufficient justification for continuing practices that have no strong evidence” (Guskey, 2015, p. 112).

## **CHAPTER II**

### **REVIEW OF THE LITERATURE**

For centuries, the purpose of education has been debated, and this debate continues today. Politicians, educators, researchers, and the general public cannot seem to find common ground on public education. Educational reforms come and go as new elections begin and end. One topic with which most agree, is the necessity of doing what is best for students.

Finding the most effective curriculum for student success has been an ever-changing process. Past philosophers and educators have debated which form of curriculum is best for students. Small (1896) believed education should be based off “the evolution of the whole person, not merely of intelligence” (p. 175). Bobbit (1912) believed that educational curriculum should:

Work up the raw material into that finished product for which it is best adapted. Applied to education this means: Educate the individual according to his capabilities. This requires that the materials of the curriculum be sufficiently various to meet the needs of every class of individuals in the community; and that the course of training and study be sufficiently flexible that the individual can be given just the things that he needs. (p. 269)

Bobbit (1924) later pointed out that education is to prepare for an adult life, not a child’s life. It is a preparation for the years after childhood. Other researchers, such as Dewey (1916), felt it was a disservice to put children on a “waiting list,” or a “probation for another life” (p. 63). Dewey, felt it was important for knowledge to become part of the child, and that the experiences a child encounters continue to evolve and grow (Kliebard, 2004). Currently, education looks somewhat different.



## **Student Learning**

Determining what a student has learned can be complicated. The educational world has many terms to describe student learning: student achievement, student learning outcomes, student success, etc. For the purpose of this paper, all of these terms will collectively be called student learning.

The executive summary from the U.S. Department of Education (USDE) *Race to the Top* (2009) program provided the following definition for student learning.

(a) For tested grades and subjects: (1) a student's score on the State's assessments under the Elementary and Secondary Education Act (ESEA); and, as appropriate, (2) other measures of student learning, such as those described in paragraph (b) of this definition, provided they are rigorous and comparable across classrooms (b). For non-tested grades and subjects: alternative measures of student learning and performance such as student scores on pre-tests and end-of-course tests; student performance on English language proficiency assessments; and other measures of student achievement that are rigorous and comparable across classrooms. (p. 14)

According to this definition, considerable focus on student learning stems from how well students do on state and classroom assessments. Teachers are responsible for ensuring student learning. USDE (2009) defines a highly effective teacher as:

A teacher whose students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth (as defined in this notice). States, LEAs, or schools must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, by student growth (as defined in this notice). Supplemental measures may include, for example, multiple observation-based assessments of teacher performance or evidence of leadership roles (which may include mentoring or leading professional learning communities) that increase the effectiveness of other teachers in the school. (p. 12)

Part of an educational leader's job is to help provide best practices in education and ensure student learning. In fact, most of the responsibility of

student success is put on educational leaders to create a culture of learning. USDE

(2009) defines a highly effective principal as:

...a principal whose students, overall and for each subgroup, achieve high student growth...other measures may include, for example, high school graduation rates; college enrollment rates; evidence of providing supportive teaching and learning conditions, strong instructional leadership, and positive family and community engagement; or evidence of attracting, developing, and retaining high numbers of effective teachers. (p. 12)

While this may appear a large task, it is important to remember a school leader's job is to do what is most appropriate for students and help them realize their potential. While the role of an educator includes teaching course content and life skills, because much of the definition of the role of an effective teacher and principal with respect to student learning revolves around assessment, it is imperative all educators become proficient in assessment practices and principals provide professional development and training so their staff can learn the necessary skills to help promote student learning (Schneider, Egan, & Julian, 2012).

### **1965 Elementary and Secondary Education Act**

Over the last 50 years, education has changed and new reforms have been introduced. In 1965, President Lyndon B. Johnson signed into law the Elementary and Secondary Education Act (ESEA) hoping to strengthen and improve educational quality and opportunities in elementary and secondary schools. A goal of the ESEA was to provide equal and appropriate learning opportunities to the nation's disadvantaged youth. Following ESEA, the federal government began providing millions of dollars to the

educational system and schools in the U.S. Between 1965 and 1980, ESEA was amended four times due to demands for higher academic standards and improved teacher preparation (Thomas & Brady, 2005).

### **A Nation at Risk**

Less than 20 years after the ESEA of 1965, the U.S. released a dissimilar report called “A Nation at Risk” in response to the accusation that the education system was broken and students in the U.S. were falling behind on international achievement tests. The publication resulted in many states increasing the academic requirements for graduation and teacher certification requirements (Thomas & Brady, 2005). The early 1990s began to shape what is known as the Standards-Based Education Reform movement. In the late 1990s, with an increased awareness of international learning, increased emphasis began to be placed on learning standards and standards-based assessments.

### **No Child Left Behind Act**

In 2001, a new reform was announced called the No Child Left Behind Act (NCLB). The NCLB Act, which reauthorizes the ESEA, puts emphasis into raising academic standards and holding state and local educational agencies accountable for student achievement. Included in NCLB was the goal for all students in the U.S. to be proficient in reading and math by 2013-2014 (NCLB, 2002).

Through NCLB, Title I regulations began requiring states to implement statewide

accountability systems covering all public schools. This accountability included annual testing and annual progress objectives ensuring all students would reach proficiency within 12 years of implementation. Data were to be broken down by poverty, ethnicity, disability, and limited English proficiency; no group is to be left behind. If schools or districts failed to make adequate yearly progress toward statewide goals, they are subject to improvement, corrective action, and restructuring to get them back on course to meet the state standards (NCLB, 2002).

Although progress is occurring from these changes in education, a review of standards-based reform confirms there are still opportunities to improve (Hamilton, Stecher, & Yuan, 2008). NCLB led educators to increase focus on student performance against the standards being taught.

### **Common Core**

In 2009, governors and education state commissioners from across the U.S. formed the Common Core State Standards (CCSS) with the goal that all U.S. children graduate from high school with the necessary skills to succeed in college career pathways and an ability to contribute to success in the economy. CCSS include learning goals identifying what students should know and be able to do at each grade level. These standards help educators focus on what students are to learn. Standards also help parents understand what is expected of their children in each grade level (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Currently, 42 States are using CCSS to educate their children.

While the majority of the States participated in creating the CCSS, the actual implementation of the standards and how they are taught and implemented is being led at the state and local levels. Although the standards are the same, educators interpret the standards differently, leading to different ideas on what is important and what is not.

The overarching goal in Common Core is to help prepare children for their future. While this is a broad goal, educators and policy makers have tried to break the goal down into more manageable steps. These steps are commonly known as grade levels. In each grade level, students are expected to learn specific items called standards. “According to Taylor, precise specifications and standards had to be established in advance in order to achieve the desired product with maximum efficiency” (as cited in Kliebard, 2004, p. 103) Standards are described as what we want students to know, and be able to do because of the education they receive in school (Guskey, 2008). These standards have been identified by policy makers as items necessary for students to learn, and put in place to prepare children for the future.

### **Race to the Top**

On February 19, 2009, President Barack Obama signed a new law called the American Recovery and Reinvestment Act (ARRA). This law is designed to stimulate the economy, support job creation, and invest in critical sectors, such as education. The ARRA helps fund educational reform by supporting and investing in innovative strategies that are most likely to lead to improved student results, long-term improvements in school system capacity, and increased productivity and effectiveness.

The money provided in the Race to the Top Fund is designed to encourage and reward States that are moving toward education innovation and reform, showing significant improvement in student outcomes, closing achievement gaps, improving high school graduation rates, and preparing students for college and career success. In addition, funds are given for implementing motivated plans in four core education reforms: (1) adopting standards and assessments that prepare students for college, the workplace, and to compete in the global economy; (2) building data systems that measure student growth and success, and inform educators how they can improve instruction; (3) recruiting, developing, rewarding, and retaining effective teachers and principals, especially in areas of high need; and (4) turning around the lowest-achieving schools.

Race to the Top rewards states that have the top plans to accelerate learning reforms in the future and demonstrate success in raising student achievement. Effective states are identified using a points-based system, in which each area is worth a specific amount of points. Rewarded states will offer models for other states to follow so effective reforms spread across the country (U.S. Department of Education, 2009).

Ultimately, the purpose of each of these reforms is to increase student success. Each reform has adapted the way educators assess students and measure student growth and learning. While reforms implemented by elected officials have a wide impact on education, student learning is measured in the classroom through the use of assessments.

### **Assessment Overview**

Throughout the course of each school year, students are administered an

assortment of assessments. The most common and frequent types of assessments are formative and happen on a daily basis. Many students are administered common or interim assessments typically measured using state standards and objectives. Last, students are administered end-of-year exams, or summative assessments, designed to measure student learning on state-mandated standards over the course of a school year. While it is ideal the formative, interim, and summative assessments all be aligned with the same standards and expectations, it is common for other factors to be measured, such as attendance, behavior, and work completion, causing confusion among educational stakeholders as to how each child is really performing.

### **Classroom assessment**

Classroom assessment is a tool educators use to gather data regarding what students know and understand (Shavelson & Towne, 2002). Classroom assessments include formative and interim assessments. These types of assessments are locally controlled and consist of informal assessment and structured techniques. These could include: student observation, questioning, tests, portfolios, reports, student self-assessment, and coursework. It allows teachers to judge student competence and learning by providing evidence gathered through different assessments (McMillan, 2012).

For most of the 20th century, the research on educational assessment focused primarily on standardized testing (Shephard, 2006). Until recently, very little research was done on classroom assessment. Through the work of Wiggins (1993) and Shephard, experts began to pay more attention to classroom assessment as a factor in effective instruction. Their research also examines learning as a process that connects existing

knowledge with new information.

While classroom teachers use data from end-of-unit summative assessments to administer grades and record student progress, formative assessment data are used in classrooms as practice to monitor students' understanding and adjust instruction to increase student learning (Cornelius, 2013). An essential part of decision making in education comes from monitoring student progress and collecting data (Hojnoski, Gischlar, & Missall, 2009). The most important reasons for schools to administer assessments are to collect data for planning instruction, evaluate the curricular program, and decide whether students are making progress toward their goals (Shapiro et al., 2012).

Today, educators have more data easily available than ever before. The push for gathering data is a result of recent research, which has shown gathering student progress data during instruction can lead to improved student achievement (Hattie, 2009). Acquiring data and making data-based decisions have become more common with the increase in accountability policies targeting student performance (Bancroft, 2010). Formative assessments occur during instruction to inform teachers of students' understanding and to guide additional instructional strategy decisions. Educational policy makers continue pushing data-informed instruction to ensure effective teaching in each classroom, ensuring assessment and accountability are at an all-time high.

### **End-of-Level or Standardized Tests**

Summative assessment occurs after instruction with the purpose of evaluating student mastery levels or demonstrating the sum of knowledge (Cornelius, 2013). The



U.S. Congress Office of Technology Assessment defined standardized tests as assessments that use uniform procedures for administration and scoring to ensure the scores from different students are comparable (as cited in Bond, 1996). As a result of federal policies, many states have created reforms to upgrade curriculum, instruction, and assessment. The federal policies have encouraged educators to implement more challenging standards, align assessments to the standards, and hold schools, administrators, and teachers responsible for helping all students become proficient in the standards (Ravitch, 1995). Some states have even gone as far to create a system of rewards and sanctions for schools and educators who do not meet the achievement targets in student performance on high-stakes tests. Because of this change, accountability policies have created incentives for schools to improve student learning (Elmore, Abelman, & Furhman, 1996). For many years, education has focused on standards in education and what students need to know to be successful. As part of the goal to ensure students are receiving the necessary skills to succeed, an increased push to identify standards every student should know has occurred. In order to ensure students are learning CCSS, policy makers instituted end-of-level high-stakes testing.

In Utah, these end-of-level high-stakes SAGE tests began during the 2013-14 school year. Each year, students take four SAGE tests: math, writing, language arts, and science. All of these tests, except the writing test, are adaptive and questions become harder as students perform better, or easier if students perform poorly. Possible scores students can receive are: below proficient, approaching proficient, proficient, and highly proficient. In addition to proficiency scores, students are given a student growth

percentile. The student growth percentile measures the amount of academic growth a student has made as a result of one year of instruction, when compared to equally performing peers. The growth calculation identifies an academic peer group of students who performed at the same level in the prior year, and produces a growth score depending on how the peer group performed on SAGE in the current year. A 50-student growth percentile means that a student made more growth than 50% of the students who had the same score the previous year.

While students' grades are not affected by the scores they receive, schools are given grades based on how well their students do on these tests. According to information compiled by the Utah State Senate (2013), "The School Grading System gives all public schools in the state of Utah a grade of A, B, C, D, or F, similar to grades given to students." The school grades are compiled using three indicators from the SAGE tests: overall student proficiency, overall student growth, and growth of nonproficient students. A fourth indicator used for high schools is high school graduation rates. As a result of the implementation of assigning school grades, politicians and the general public can look at school grades and compare the testing results for all public schools in the state of Utah.

Utah is not the only state that has made changes; an increasing number of standards-based tests are administered to students across the country. For several decades, U.S. students' standardized test scores have been compared with other countries' test scores. In a push to be the best, an emphasis on ensuring students know the learning standards, and are being tested on those standards, has occurred.

Clearly, testing has its place in education. If used correctly, assessment can have a

positive effect on student achievement and student learning. Politicians have used high-stakes tests as a way to leverage the problem of low student achievement and underperforming schools (Smith, Heinecke, & Noble, 1999). Educational leaders have the important task of ensuring students receive the best and most appropriate education possible. It is important teachers are trained on proper assessment techniques to ensure student learning. Educational leaders also have the task of working with legislators and school boards to ensure standardized testing is accomplishing the task it is set out to do.

School districts use data from summative, high-stakes state assessments to report student learning. Much of the focus and reasoning behind high-stakes testing is for teacher and school accountability. While very little pressure is put on students for high-stakes tests—because they do not affect student grades or graduation—teachers, schools, and districts feel pressure as they are held accountable. Naturally, this leads to an increased focus on learning standards to prepare students for the tests.

### **Grading**

Teachers' decisions on grades and student achievement can have a long-lasting impact for students. "Grading is the process of summing up student achievement with marks or symbols. Teachers grade individual assessments and assignments, and they also combine those grades into one summary mark or symbol for a report period" (Brookhart, 2012, p. 257). There are many different purposes for grading. Researchers have categorized teacher and school leader responses into six major categories (Airasian, 2001; Feldmesser, 1971; Frisbie & Waltman, 1992; Guskey & Bailey, 2001; Linn, 1983).

1. Communicate information about student achievement to parents and others.
2. Provide information and feedback to students for self-evaluation.
3. Select, identify, or group students for specific educational paths and programs. This could include higher level AP and honors classes or possibly remedial and special education classes.
4. Provide incentive and motivation for students to learn.
5. Evaluate the effectiveness of instructional programs. Are students learning and is the content valuable?
6. Provide evidence of students' lack of effort or inappropriate responsibility.

Educators generally agree each of the six purposes is important and legitimate, but they seldom agree which one is most important. The fact that educators cannot agree on which purpose is the most important helps illustrate why there are so many different grading practices in education.

### **History of Grading**

Student grades have long been in use to gauge student ability. In the 1800s, a method was needed to determine which students were doing well in school. Educators began using percentage grading to assess learning and determine the percent of points earned on assignments and tests. In the early 1900s, the letter grade became popular as a method to categorize percentages. The A-F scale was implemented using plusses and minuses as a way to differentiate even further. End-of-term grades were then calculated and students were given a grade point average (GPA). This is the grading scale most commonly used in schools today. Typically, letter grades are factored using percentage grading from 0% to 100% based on the total number of points earned on assessments.

Starch and Elliot (1912) state,

The reliability of the school's estimate of the accomplishment and progress of pupils is of large practical importance. For, after all, the marks or grades attached to a pupil's work are the tangible measure of the result of his attainments, and constitute the chief basis for the determination of essential administrative problems of the school, such as transfer, promotion, retardation, elimination, and admission to higher institutions, to say nothing of the problem of the influence of these marks upon the moral attitude of the pupil toward the school, education, and even life. The recent studies of grades have emphatically directed our attention to the wide variation and utter absence of standards in the assignment of values. (p. 442)

While this grading system has been around for many years, it does have some flaws. Depending on each teacher, this grading system can serve a variety of purposes. Teachers have been known to use behavior, responsibility, and attitude to help determine a child's grade. When specific and common measures are not used in grading, it is up to the teacher's discretion to assign grades, and almost anything can be assessed and counted toward a grade (O'Connor & Wormeli, 2011). In an effort to separate what a student has learned and can demonstrate and a student's behavior, many schools have created a secondary, or citizenship grade to assess a student's behavior in each specific class. While the citizenship grade is in place, many educators still fail to separate the behavior and citizenship of the student from the letter grade, designed to measure what standards and content students have learned.

### **Norm-Referenced Grading**

Concerns with the traditional methods of assigning student grades led to a different type of grading called norm-referenced grading or "bell curve" grading. This grading method is used to compare students to each other, allowing a small percentage of students to score well, most to score average, and small percentage to score poorly. A

concern with norm-referenced grading is it limits the number of students who can earn a top grade (Raymond, 2013).

Norm-referenced grading assumes the average student is going to earn a C and the majority of students fall into average range, causing a negative reaction since one of the purposes of teaching is to improve the knowledge base of all students. Once teaching has taken place, it is natural to assume all students should rank above the average in knowledge. Norm-referenced grading curves scores and ensures students are competing against each other, rather than competing against themselves and the standard (Raymond, 2013). Teachers who have implemented norm-referenced grading set scales so only a percentage of students will be ranked in the top tier of the grading scale. Even though the lowest students may have achieved at high levels, they will be lumped into the lowest grade level (Guskey, 2001). This leads students to compete against each other and creates a reluctance to help each other and collaborate in learning. Instead of helping one another and learning together, learning can become a game of winners and losers (Haladyna, 1999). One of the main reasons educators use norm-referenced grading is for convenience; however, there is no educational justification for grading on a curve (Biggs, 1999). Norm-referenced grading has been criticized because the focus is on competition, rather than assessing skills and content (Bond, 1996). Educators who grew concerned with norm-referenced grading began to look for a more effective way to assess student learning.

### **Mastery-Based Grading**

Bloom (1984) recognized a concern with comparing students against one another,

rather than against learning targets. Mastery-based grading allows students to pass or move on once they have reached a pre-specified level. Different amounts of time are allotted to students based on needs and the time they require to reach mastery. Mastery learning requires each student to achieve established standards of performance. The focus is on how well students are doing against the standards and objective, not against each other. Students are required to learn, and if necessary relearn, until mastery is demonstrated. Because students learn at different levels and times, enrichment activities are needed for students who demonstrate mastery earlier than others. In mastery learning, clear standards and objectives must be identified, assessments must be created to determine mastery, and grading incentives must be established to encourage students to learn beyond initial mastery (Lalley, & Gentile, 2009).

Mastery-based grading led into an experiment to eliminate the traditional letter grading scale and create two levels of grades: pass/fail. These two indicators were used to determine if a student met mastery or not. Some educators felt mastery-based learning was effective to help individualize education, but it was difficult for the teachers to track student progress and prepare what students needed next. Several areas of the U.S. have recently re-introduced mastery-based grading because improved technology and computer based learning programs have made it easier on educators to track students.

### **Criterion-Based Grading**

A criterion is “a distinguishing property or characteristic of anything by which its quality can be judged or estimated, or by which a decision or classification can be made.” Its purpose is “to describe, clarify, and communicate requirements, to contextualize and

fine-tune expectations; to facilitate the substantiation of judgments; to safeguard against subjectivity and bias; to ensure fairness; and to provide a defensible framework for assessing” (Scarino, 2005, p. 9). In a criterion-referenced grading system, student performance is measured against a fixed set of learning criteria. Students are scored and evaluated against a set number of points or the percentage system. The focus is on how well a student has learned a specific body of knowledge rather than a comparison of students. One of the goals of criterion-referenced grading is to help all students learn enough to meet the standards and pass each assessment. Because comparison to other students is not used, everybody works toward the same goal of meeting and exceeding the predetermined performance levels. Criterion-based assessments may include multiple-choice, true-false, or open-ended questions. Many assessments include a variety of these types of questions to determine if the students have adequately learned the required content. Criterion-referenced tests are the most widely used type of test in the U.S. public school setting. Through this grading method, results are obtained to determine to what extent the learner has achieved the course objectives and help educators know if the learner is ready to move on to the next level of instruction (Smith, 1973). Standards-based grading measures how well a student knows the predetermined standard and can demonstrate learning on levels rated as advanced or proficient.

With criterion-referenced tests, educators are able to determine how well students are doing compared to pre-determined performance levels in the curriculum (Bond, 1996). Seat time and age are not factors in advancing to the next level. Students are able to work at their own pace and are evaluated on their performance and how they apply



their knowledge of the subject matter. Students are evaluated on performance and application of what they are learning, irrespective of how their peers are doing. Through criterion-based learning, educators are able to differentiate learning for individual students and assess skills in a myriad of ways (Sturgis, Patrick, & Pittenger, 2011). Though educators have experimented with different types of grading practices over the years, most educators continue to resort back to the most common grading methods and traditional grading practices.

### **Problems with Traditional Grading Practices**

Traditional grading practices are those practices most used by educators and have been around for over a century. This method of grading provides minimal information about what a student has learned (Marzano & Heflebower, 2011). In some cases, teachers have been known to inflate grades to improve teacher feedback from students (Griffin, Hilton, Plummer, & Barret, 2014). According to Welsh, D'Agostino, and Kaniskan (2013), "The grading practices teachers use may jeopardize the reliability of grades and therefore weaken the link between grades and academic achievement" (p. 27).

Teachers' grading practices are not only different from the recommendations of specialists; they also vary considerably from teacher to teacher and are occasionally inconsistent within a teacher's own classroom (Brookhart, 1993; Friedman & Trog, 1998). In some classrooms, certain assignments, assessments, and behaviors have been known to carry different weights toward a child's grade (Resh, 2009). This thorough examination of classroom grading practices is revealing as it provides an example of how various policies and principles are translated into practice. It also provides awareness into

understanding the various focal points, personal views, and administrative restrictions that may promote or oppose the ideas executed within a standards-based context.

Through their studies, Starch and Elliot (1912) found when different teachers are given the same assignment to grade, student scores differ from teacher to teacher. There is unreliability in the standards teachers look at and the weight by which assessments are scored. If each teacher is using different techniques in grading, determining what a grade means and if a student has obtained the desired academic achievement and learned the standards of each course can become more difficult.

Brookhart (2004) reviewed a set of 19 studies regarding grading practices and found four main findings. First, most teachers try hard to be fair to students by informing them up front what the components of their grade will be. Next, tests and other achievement measures make up the majority of the grade, but ability and effort are commonly considered. Third, grading practices differ by grade level. Elementary teachers tend to use more informal evidence and observation to grade their students while secondary teachers use tests, quizzes, and other written activities to comprise the majority of a student's grade. Last, there is wide variation from teacher to teacher in grading practices. Each teacher perceives the meaning and purpose of grades differently. The variation in grading practices is alarming because it raises concerns about the validity of grading and what those grades are supposed to measure.

The meaning of grades varies based on the teacher, department, school, and state a student receives their education. "Validity is in question when the construct to be measured is not purely achievement but rather some mix of achievement and non-

achievement factors...it is also in question when grades mean different things in different schools or subjects, in different teachers' classes, and for different types of students" (Brookhart, 2012, p. 260). Many teachers seem to resist the idea of grading on achievement alone because they do not understand the negative effects of mixing methods when grading. In order for teachers to change this mindset, Cross and Frary (1999) recommend a solution of providing teachers with professional development about the effects of their grading practices and continued research on grading practices, including training in standards-based grading.

The meaning of grades and grading practices have been shown to be perceived differently by parents compared to teachers (Waltman & Frisbie, 1994). Inconsistencies in the way teachers grade can be found in many circumstances. In a study by Feldman, Kropf, and Allibrandi (1998) results showed most teachers believed grades should reflect learning and achievement, yet many of those same teachers included points for completing assignments, doing extra credit, and studying as part of a student's grade. While teachers reported interest in grading on learning, they were in fact grading behavior unrelated to student learning. In a study by Randall and Engelhard (2010), survey results show student grades increase as behavior improves, regardless of academic ability.

Grades are open to multiple interpretations and do not have a comprehensive meaning for all stakeholders. Because of this, it is suggested conventional grading practices should change to better support student autonomy and improved student learning (Thomas & Oldfather, 1997).

Standards-focused programs like the Common Core typically have a centrally developed curriculum, common reporting mechanisms, and large-scale assessment programs. Evaluation of the impact of standards-based movements has shown some gains in student achievement (Hanushek & Raymond, 2004), but currently there is very little evidence in which standards-based systems have actually improved the quality or consistency of grading practices in the classroom.

Grading practices tend to focus specifically on students' end of term grades as determined by the teacher for report cards. Effective grading should be based on varied and multiple sources of evidence. While the use of these sources of evidence make it more likely to understand what a child can do, it also makes the grading process more difficult and less straightforward for teachers (National Council of Teachers of Mathematics, 2000). Many of the concerns raised in the past about grading, such as the misinterpretations of report cards, including nonacademic factors in grades, and the unreliability of percentage systems, continue to resurface (Brookhart, 2004).

In education, much of assessing whether or not students have learned the required standards is shown through students' grades and test scores. While both grades and test scores can show what students know, they are compiled differently and show dissimilar data in regards to student learning. Best practice encourages educators to look at their current grading practices and make the adjustments needed to maximize student learning.

Although focus on standards has increased, grading practices have mostly remained the same. One issue of using averages is students and parents are not able to determine which standards have been met and which standards need to be re-emphasized

and retaught. As a result, many educators have begun implementing a grading system called standards-based grading.

### Standards-Based Grading

Standards-based grading is described as a grading system in which students are evaluated based on their proficiency in meeting clearly articulated standards (Tomlinson & McTighe, 2006). The idea behind standards-based grading is to focus on the larger outcome of learning, rather than attendance, work completion, and how many points a student has accumulated from any number of assignments (Table 1). Standards-based

Table 1

#### *Common Differences Between Traditional Grading and Standards-Based Grading*

Traditional grading	Standards-based grading
Penalties for late work	No penalties for late work
Including nonacademic factors into the grade	Only factoring academic factors into grades
Everything goes in the grade book, regardless of purpose	Selected assessments are used for grading purposes
Includes every score regardless of when it was collected.	Emphasize the most recent evidence of learning
Based on the percentage system	Criterion or proficiency-based.
Grade deductions for retaking assessments	Students are allowed multiple opportunities to retake assessments without penalty
Homework counts heavily towards the grade	Homework is for learning practice.
Learning standards are combined to show one grade	Learning standards are separated to show proficiency levels on each standard
Zeros are counted in a grade	Zeros are removed and counted as incomplete
Students are unsure of learning targets	Learning targets are made available to students before the assessment is given

Guskey, 2015; O'Connor, 2002; Schimmer, 2016.

grading encourages teachers to implement more measurable methods when evaluating student learning and progress. It also encourages students to meet learning standards and become proficient in each respective standard (Iamarino, 2014). Some grading practices compare students against other students. Using standards-based grading, students are evaluated by the standard which can be differentiated by levels of quality (Guskey, 2001).

Research on standards-based grading is still in its infancy (Brookhart, 2012). Currently, there is more professional literature (e.g., Brookhart, 2012; Guskey & Bailey, 2010; McTighe & O'Connor, 2005) related to standards-based grading than research literature. Most of the current evidence regarding standards-based grading is anecdotal (Brookhart, 2012).

Standards-based grading enables educators to grade students based on how well they perform at each standard. Instead of averaging scores for a set amount of time, students are graded based on how they demonstrate knowledge in each respective standard (Hooper & Cowell, 2014). Clearly describing the expected standards of performance is important so students know what they are expected to learn (Buckles et al., 2001). Using this method provides the opportunity for educators, students, and parents to know the standards and identify where students need more support and learning interventions. Since standards are used to guide the curriculum, it makes sense to assess on the mastery of those same standards.

Standards-based grading has been developed and implemented in many elementary schools. The implementation of standards-based grading in secondary schools

has been slow because of the wide variety of classes and the many standards taught in each subject. In addition, letter grades are still the norm when applying for post high school education.

McMunn, Schenck, and McColskey (2003) studied the professional development in standards-based grading practices in a Florida school district. Results indicated teachers felt the provided professional development allowed them to be more focused on an achievement-based form of grading. However, observations and classroom artifacts showed this was only true for a portion of the teachers. It was determined educators need more intensive support and professional development before undertaking a change in grading practices. While standards-based grading research is still in its infancy, many educators are implementing it within their classrooms and schools because they feel it is most appropriate for student learning and growth.

Standards-based grading does not measure student learning the same way traditional grading does. Grading experts argue standards-based grading provides a more accurate description of what a student knows and can do (Guskey, 2015; Marzano, 2006; O'Conner, 2011; Schimmer, 2016). In order for educators to truly implement standards-based grading effectively, a change from traditional grading mindset must take place.

Changing the way educators teach and having a standards-based mindset means educators redefine accountability, repurpose the role of homework, and give students full credit for what they know (Schimmer, 2016). CCSS provide an opportunity for schools to move toward standards-based grading. Educators can base grades on common reference points, or standards, making it possible to develop common grading policies. Creating

common grading policies will result in less variability among teachers and schools (Canady, Canady, & Meek, 2017).

### **Preparing Students for the Real World and Holding Them Accountable**

Often teachers feel that if they allow students to hand work in late, they are not holding students accountable for their work. The argument can be had that students are not being held accountable if they are simply given a zero for not completing an assignment. If a child does not clean their room, is giving them a zero for the day holding them accountable or is giving them more time and a second chance to clean their room holding them accountable? In the business world, the most productive growth environments are those where employees see constructive feedback as a source of empowerment and can make adjustments to their work based on that feedback (Walker, 2002). Educators need to take caution in how they are describing the real world to students. Often times, the real world educators describe is, in fact, not the real world at all. Businesses are trying to create productive and supportive work environments, personalized to get the most out of their employees. Many educators feel justified in giving academic penalties for work handed in late, despite the fact there are no studies that support the use of assigning low grades as an effective punishment (Guskey, 2004).

One hundred years ago, education was preparing students to enter a workforce full of industrial jobs that required a certain skill set. Modern technology has made most of those jobs obsolete. Preparing students for the current workforce requires a different approach. Guskey (2015) stated,



Educators must do their best to help every student develop the advanced knowledge and 21<sup>st</sup> century skills necessary to enter college or be ready to begin a career that requires creativity, collaboration, advanced reasoning, and problem solving. (p. 4).

These 21<sup>st</sup> century skills can be learned through effective classroom discussions, group projects, and blended and personalized learning. Educators should train students to be thinkers and self-starters, not robots to complete mundane and repetitive tasks.

Most tests individuals take outside of K-12 education, including the ACT, SAT, state law bar exams, and graduate school entrance exams, provide allowances for participants to take the test as many times as they wish. The highest score of each test is recorded and kept. Penalties are not given for retakes. Standards-based grading practices require teachers to organize evidence of student learning by standards and factor out all of the nonachievement qualities (Schimmer, 2016).

### **Grading Behavior, Attendance, and Citizenship**

A common concern amongst educators who learn about standards-based grading is the notion that behavior, attendance, work completion, attitude, and overall classroom citizenship should not be included in the letter grade. A typical response from teachers is the need to teach students to be accountable and responsible for their actions. For the most part, students want to achieve high grades and teachers have been able to get the best behavior out of most students by dangling a grade in front of them that includes student behavior and other items not related to the learning standards. Standards-based grading proponents have suggested creating a second grade that examines a student's behavior and overall citizenship. By having multiple grades, effort, attendance, behavior,

citizenship, and other important life skills can still be assessed, but they would be separate from assessments and content learning standards (Guskey, 2002; Stiggins & Chappius, 2011). Many schools across the world already have secondary grade to determine a student's citizenship in each class. Educators already recognize classroom citizenship is an important trait for students to learn, but they also need to recognize citizenship is different from knowledge and skills and therefore, must be assessed and graded differently (Guskey, 2015).

### **Identified and Clear Learning Targets**

Learning standards are often broad and nearly impossible to cover over the span of a school year. Educators are tasked with determining which of the standards are most important and needed for the next school year or next phase of life. Teacher collaborative teams should identify clear learning targets to ensure students have access to the same learning, regardless of who is teaching the class (Marzano, 2003). Teachers create pacing guides and curriculum maps and identify which standards students need to know. When determining standards, it is important the standards have lasting value to the student, apply to multiple academic disciplines, and are necessary for the next level of instruction (Reeves, 2004).

While it is imperative educators become familiar with the standards to identify clear learning targets, it is also important students are aware of what they should be learning. To do this, educators are encouraged to create scoring guides or rubrics. These documents provide a description that clearly differentiate levels of performance to determine proficiency (Reeves, 2004). "Teachers are most effective in helping all

students in helping all students learn when they are clear regarding exactly what their students must know and be able to do as a result of the course, grade level, or unit of instruction” (Dufour, Dufour, Eaker, & Many, 2010, p. 70). As teachers identify and deepen their understanding of learning standards, assessing and grading the knowledge students demonstrate becomes easier.

### **Counting New Evidence and Getting Rid of the Average**

In traditional grading methods, grades are often averaged with what students currently know and what students used to not know (Schimmer, 2016). Teachers will often create grading formulas that too often do not reflect actual levels of proficiency. As students learn and gain new levels of proficiency, older evidence of learning becomes invalid (Guskey, 2015). When a teacher assigns a grade, they are looking to be as accurate as possible to reflect what a student knows. Using current knowledge over old knowledge is a more accurate way of reporting. If a parent were to grade their 5-year-old on how well he or she walks, would it be accurate to assign a grade that includes data from the transition stage of crawling and walking? Most 5-year-old children walk very well, so assigning a grade that includes the beginning stages of walking would be inaccurate. In the same regard, including scores at the beginning stages of learning a standard is inaccurate as well. Student grades should not be effected because they are in the beginning stages of learning. If a student can demonstrate he or she is continuing to learn and progress, the grades should be a representation of the most recent knowledge a student has. Averaging scores for the same standard rewards students who learn quickly

and punishes students who take longer to grasp concepts (Schimmer, 2016).

### **Using Homework Practice as A Grade**

A common practice in education is to assign and grade homework. Educators feel student learning will increase if students are spending extra time at home learning the content. While this may be true, educators also value homework because it teaches students responsibility and a good work ethic (Dueck, 2014). Grading homework and factoring it into students' grades does have its disadvantages. Completing homework becomes more of how much time a student spends on the homework, rather than what the student knows. Chris Van Bergeyk (as cited in Dueck, 2014) indicated:

Hard-working students get extra marks through completing homework without deepening their understanding. It's akin to bonus marks—students may be tempted to spend two hours on homework to make it look great, and to subsequently receive the extra 10 percent for completion. Homework really acts as grading cushion for many students, especially hard-working ones. (p. 46)

Because homework is most often done without adult supervision, the opportunity to cheat becomes more likely. Homework can be a tool to increase learning, but it also increases the learning gap for at-risk students. According to Jensen (2009), at-risk students living in poverty:

- Are more likely to live in a crowded home.
- Inherit low self-esteem.
- Own fewer books and watch more TV than their peers.
- Inherit negative views of school.
- Have a 50% chance of dealing with evictions, utility disconnections, overcrowding, or lack of a refrigerator.
- Have mentally adapted to suboptimal conditions.

- Are tardy and absent more often than their peers.
- Have more physical altercations more often than their peers.
- Are more likely than their peers to experience physical punishment.

All of these factors would limit a student from being able to complete their homework. At-risk students have different responsibilities than their peers that may include: having a job, babysitting while their parents are at work, cooking meals, cleaning, and other tasks that would prevent a child from working on their homework.

Michell Nikisch (as cited in Dueck, 2014), an educator in Washington State, stated:

I stopped grading homework the day I found out that one of my students had to wait until her father came home from work so that she could stand out in the gravel driveway and use the headlights from his truck to complete her homework. It was not fair of me to ask a student to do that. That day, I realized I could never ask a student to do that. That day, I realized I could never again ask a student to stand outside in the elements just for a homework grade. (p. 53)

Assigning grading penalties for things beyond a students' control is academically inaccurate and morally wrong (Dueck, 2014). Vatterott (2009) touches on the homework dilemma teachers face, "The attempts of researchers to answer that basic question (does homework improve achievement) have led to conclusions that are inconsistent at best and contradictory at worst" (p. 58). Despite some of the research against homework, educators know support and learning at home are necessary to maximize learning. Providing practice at home that is age and time appropriate can produce positive results (Marzano, Pickering, & Pollock, 2001). Thus, educators using standards-based grading view homework as an opportunity to practice and improve, rather than a form of assessment to be counted on the grade.

## Use of Zeroes

When a student is absent for a day or forgets to turn in an assignment, the teacher will often assign a grade of zero for that assignment. Reeves (2010) indicated,

On a four-point scale, where A = 4, B = 3, and so on the zero is accurate, because the difference between the A, B, C, D, and F are all equal - one point. But assigning a zero on a 100-point scale is a math error; it implies a 60-point difference between the D and F, while the other differences are typically about 10 points. It makes missing a single assignment the 'academic death penalty.' It's not just unfair - it is not mathematically accurate. (p.78)

To improve the accuracy of grades when using a 100-point scale, instead of giving scores of zeroes for work that is not turned in, educators should consider using incompletes and interventions to motivate the students to complete the work. Giving a student a zero will skew the grade and make it inaccurate (Dueck, 2014).

Canady (n.d.) gave a real world example of how a zero can skew the numbers:

Suppose we are developing a brochure for the Chamber of Commerce and our goal is to report a reliable average temperature for a week in September in Charlottesville, Virginia. The temperatures for the week, taken at noon on Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday, were 92, 91, 90, 80, 84, 85, and 82, yielding a total of 604. If we divide 604 by 7, we get an average of 86. Let's suppose, however, that we did not have a temperature reading for Wednesday, and we put a zero for Wednesday and still divide by 7; then we get an average of 75. Suppose we treated Wednesday simply as missing data and divided the total by 6 and not 7, then we would get an average of 87. Which summative average is the most reliable average temperature to report to the Chamber of Commerce brochure: 75 or 87? We can see that the zero poses a problem in averaging temperatures for the week. (p. 6)

Assessment researcher, Marzano (2006) has proposed alternative scales that are more balanced and fair. Marzano uses a 0- to 4-point scale ensuring the scale is balanced. A score of zero shows a student has shown no understanding or demonstration of the learning standard even when provided help. A 1 on the Marzano scale shows that, with

help, a student demonstrates partial understanding of the simplest details and processes of the learning standard. A score of 2 shows no major errors or omissions on the simpler details, but major errors on the more advanced details and processes. A 3 shows no major errors regarding the learning standard. Last, a 4 shows a student has earned a 3 but can also provide in-depth inferences and apply what was taught. Although a zero is used in the Marzano scale, it does not skew the grade in a way that prevents the student from recovering.

### **Reassessment**

The idea of allowing a student a second opportunity to test is often out of the question when using traditional grading methods. Traditional classes and methods of grading argue that tests cannot be revisited. The traditional way of thinking argues that students have one chance in life and if they do not succeed then they should learn from their failure and work harder next time. While educators like to argue, you do not have second chances in real life, that is just not true. Rshaid (2011) said,

Our education system is supposed to prepare students for real life, and it is easy to see that this artificial assessment model has little resemblance to reality. In real life there are almost no once chance do-or-die scenarios, and whenever anybody has to demonstrate proficiency in any field, the timing for demonstrating that competency is chosen by the candidate rather than being an arbitrary date set in stone. (p. 26)

Even in the “real-world” of education students can take the ACT and SAT as many times as they wish. Law students are able to take the bar exam as many times as they need to pass. Certified personal accountants (CPAs) can take the CPA test until a passing score is obtained. It is important educators remember that a test is a snapshot in

time. Variables such as not having breakfast, running late, getting in an argument with a parent, or a boyfriend breaking up with them can affect the results of a test. Outside factors can negatively affect all students (Dueck, 2014). The real world allows more than one chance, and teachers who do not allow second chances are punishing their students. Canady et al. (2017) said, “It is far more realistic to teach students that mistakes do happen and that they must continue trying after failure” (p. 52). Teaching students to not give up and to keep persevering is an important life skill to learn.

However, reassessing is much more than a second chance, it is an opportunity for students to demonstrate what a teacher already suspects they have learned. A key part of allowing reassessments is ensuring students have learned something new. It does no good to have a student continually reassess if they have not gained any new knowledge. As students learn, grow, and gain more knowledge, it is important teachers allow students to demonstrate that growth (Schimmer, 2016). Otherwise, there is little value in continuing to learn.

In order for reassessing to be effective, educators teaching common classes should have common assessments, common curricular pacing, time for teacher collaboration and data analysis, and strong teacher and administrative leadership (Cox, 2011). Hirst (1973) stated,

Its greatest promise seems to be its ability to pace the student in a meaningful relationship with the domain of knowledge so that his progress can be measured based entirely on his individual efforts.... What is more humane than letting the learner know, in advance, what he is expected to know, under what conditions he will be expected to know, under what conditions he will be expected to demonstrate his knowledge, and the level or degree of acceptable performance he is to achieve. (pp. 50-52)



### **Grades Versus High-Stakes Testing Scores**

Some research has already been completed examining the relationship between students' grades and their high-stakes test scores. Generally, these studies have found a moderate association between grades and test scores. Discrepancies are often credited to teachers using nonacademic factors when assigning student grades (Brennan, Kim, Wenz-Gross, & Siperstein, 2001; Martinez, Stecher, & Borko, 2009; Willingham, Pollack, & Lewis, 2002).

SAGE tests, the current end-of-level assessment administered in Utah, at the end of each school year, focus on the standards taught in math, English, and science. Students are assessed on the content standards for those subjects on the SAGE test, yet many teachers still use traditional grading practices in the classroom not focused on the standards. Traditionally, students are graded largely on averages compiled over a certain amount of time. An issue with using averages is that parents, teachers, and students are unclear about what standards have been met and what standards need to be re-taught. Also, the traditional grading system can serve a variety of purposes. Teachers have been known to use behavior, responsibility, and attitude to help determine a child's grade. When specific measures are not used, teacher discretion determines a child's grade causing it to be unreliable from class to class. As a result, many teachers have begun to implement standards-based grading with the intent to focus on content standards.

Welsh et al. (2013) researched how standards-based grades related to student achievement on state tests for third and fifth graders. Results show there is a wide variation among teachers in how closely their students' grades matched the achievement

standards on the state tests. In addition, the researchers coded information from teacher interviews regarding how closely they followed recommended grading practices. Results show the more closely teachers followed the recommended grading practices, the higher the correlation between students' grades and test scores. Currently, there is a lack of literature researching the correlation between standards-based grades and end-of-level test scores in secondary schools. While researchers believe standards-based grading positively helps student learning, no major research has been completed demonstrating that claim (Marzano, 2011). Brookhart (2012) believes studies related to the claims that standards-based grading increases student learning would be helpful to add to existing research.

### **Conclusion**

Grading reforms have come and gone. While some grading practices might catch on with certain teachers, for the most part, educators revert back to traditional grading methods and the grading practices that were used when they were in school (Guskey, 2015). Every year, educators are tasked with different changes put on them by politicians and district and school leaders. Because of all the changes teachers have to implement, many of them grow weary of changing something else. If real changes in educational grading practices are going to happen, research must be provided to determine their effectiveness.

## **CHAPTER III**

### **METHODS**

The purpose of this study was to evaluate the effectiveness of determining student learning through the use of standards-based grading in secondary classrooms and the correlation those grades have with end-of-level test scores. The use of standards-based grading by teachers in the study was an adjustment of grading practices from traditional grading. Adjusting grading practices to support student learning was reinforced by Scriffiny (2008) when she commented, “If your grading system doesn't guide students toward excellence, it's time for something completely different” (p. 70). As educators reflect on their grading practices, it is important they determine if grading practices accurately reflect student learning. Self-reflection in grading is an important practice of educators seeking to do what is best for their students and help students reach their goals.

Major goals in education include improving student achievement, closing achievement gaps, and preparing for student success in college and career (USDE, 2009). For many years, education has focused on standards in education and what students need to know to be successful (Marzano, 2003). Standards are described as what we want students to know and be able to do because of the education they receive in school (Guskey, 2008). One of the reasons for standards-based grading is to ensure consistent areas of instructional focus that prepares children for the future.

Although the focus on learning standards has been widely present, grading practices have received modest attention and changed very little. Students continue to be graded largely on averages compiled over a certain amount of time. Unfortunately,

grading averages do not clearly communicate to students and parents which standards have been mastered and which standards may need further instruction or additional practice. Furthermore, the traditional grading system can serve a variety of purposes. For example, grades assigned by teachers commonly include indicators unrelated to learning, such as student behavior, participation, and homework completion, while the end-of-level exams given to students are scored based off how well each student demonstrates competency of the standards.

### **Research Questions**

For the focus of this study, the research sought to determine if standards-based grading is an effective way to assess student learning in comparison with end-of-level test scores. Additionally, the researcher examined the effects of standards-based grading on increasing learning as well as the impact of teachers' years of experience on student grades and outcomes. In order to address the areas of focus identified, the following four research questions guided this study.

1. Is there a positive correlation between standards-based grading term grades and end-of-level SAGE test scores?
2. To what extent does the standards-based grading environment increase or decrease student learning, in comparison with traditional grading, as measured by the student growth percentage on end-of-level SAGE test scores?
3. Is there a difference in the correlation of standards-based grades and end-of-level SAGE test scores between math, language arts, and science?
4. Does the correlation between standards-based grades and end-of-level test scores vary based on teachers' years of teaching experience?

### **Research Design**

For this study, the researcher used a quantitative, correlational research design to examine the relationship between grades (traditional grading and standards-based grading) and end-of-level SAGE test scores in mathematics, language arts, and science. A correlational design was most appropriate for this study because the researcher was looking to examine the relationship between two or more variables. Correlational research can be an effective tool in discovering the relationships between variables, establishing the level of these relationships, and providing important information for conducting higher-level research (Fraenkel & Wallen, 2006).

### **Setting**

The setting for this research study included a large, suburban school district in the western U.S. The school district consists of 89 schools with over 70,000 students. Included in the 89 schools are 17 junior high schools and 10 high schools. The junior high schools consist of students in grades 7-9, while high schools include students in grades 10-12. Only teachers from junior high schools and high schools participated in the study.

### **Participants**

Participants in this research study were assigned to two groups: standards-based grading and traditional grading. There were a total of 90 participants: 45 teachers in the standards-based grading cohort and 45 teachers in the traditional grading group. The

standards-based grading cohort was an established cohort of secondary teachers engaged in a multi-year professional learning experience focused on developing standards-based grading practices. The traditional grading group was comprised of teachers not involved in the professional learning cohort.

### **Standards-Based Grading Cohort**

The first group of participants in this study were junior high and high school math, science, or language arts teachers participating in a standards-based grading professional learning cohort. All junior high and high school math, science, and language arts teachers in the school district were invited to participate via email in the standards-based grading cohort. Forty-five teachers (Table 2) volunteered to participate in the standards-based grading professional learning cohort.

The teachers who volunteered for the standards-based grading cohort read the book *Grading Smarter, Not Harder* by Myron Dueck. The book was used to facilitate discussions to build agreement on the foundation and purpose of the grading process and what grades mean. The standards-based grading cohort met together during the 2015-16

Table 2

#### *Standards-Based Grading Cohort and Traditional Grading Group Demographics*

Group	N	Gender	Years of experience	Subject taught
Standards-based cohort	45	Female: 38 Male: 7	0-5: 21 6-10: 12 11-15: 8 16+: 4	Language Arts: 19 Math: 14 Science: 12
Traditional group	45	Female: 35 Male: 10	0-5: 19 6-10: 10 11-15: 7 16+: 9	Language Arts: 19 Math: 14 Science: 12

school year every month for 5 months and every other month for 4 months with district curriculum specialists. During the meetings, implementation strategies were discussed, helpful tips were given through the use of articles, books, and presentations; and time was spent discussing what was working effectively and areas that were of concern. Classroom implementation of standards-based grading practices continued during the 2016-17 school year.

### **Traditional Grading Group**

In order to understand the effect of standards-based grading practices, a second group of teachers were selected for the study. The second group of participants (Table 2) were selected from the remaining junior high and high school math, science, or language arts teachers who were not participating in the standards-based cohort which was approximately 759 teachers. This group was further refined by conducting a demographic match to the teachers involved in the standards-based cohort. The match was used to increase the validity of the traditional grading group, therefore, participant selection data controls were instituted to produce a similarly matched group. Specifically, the filters used included: school demographics, grade level, subject matter, and years of experience in teaching. Applying such filters improved the likelihood of similar cases between the two groups rather than a random group from the general population. Teachers in the traditional grading group did not receive any grading interventions or professional development from the school district. The teachers in this group were unaware of their participation in this study.

### **Data Sources**

The researcher collected two primary sources of data for this study: 2016-17 end of term grades and Spring 2017 end-of-level SAGE test scores. The grades used for this study were collected from the school district's data bank at the end of the 2016-17 school year for math, science, and language arts classes. End-of-level SAGE test data included the scale score and growth percentile for each student in the data set.

The data collected included end of term grades and end-of-level SAGE test scores for students ( $n = 8,129$ ) who were enrolled in the classes of teachers in the standards-based cohort and teachers in the traditional grading group of. Students ranged in age from 12-18 and attended grades 7-12. While all students lived in the same school district, they attended 24 different schools, consisting of 8 high schools and 16 junior high schools. The students from participating schools have 15.89% minority rate and 20% of them come from low income families. Student information was de-identified and the school district's data and assessment team used ex post facto to pull student grades and SAGE test scores.

### **Data Collection Procedures**

Data collection consisted of two phases, the majority of which came from the school district's assessment and evaluation department. In phase one, the researcher collected data to ensure teachers in the standards-based grading cohort were implementing standards-based grading as prescribed in their training. Then, in phase two the researcher collected grades and end-of-level SAGE test scores for students enrolled in



the participants' classes. The following sections provides a detailed description of each phase.

### **Phase One**

During the 2016-17 school year, the standards-based grading teachers participated in a survey to determine their level of implementation of standards-based practices. The survey was developed by a district committee consisting of school administrators, curriculum supervisors, and assessment technicians focused on analyzing the standards-based grading in the cohort classrooms. The committee created a list of potential questions to survey the standards-based grading teachers. The list of questions was reviewed using Lawshe's (1975) method by a diverse group of educators including: a state of Utah assistant superintendent, district curriculum director, district curriculum supervisors for math, science, and language arts, the district assessment director, and the researcher. Questions that did not meet the validation method were excluded from the study. In the end, all questions included in the survey (Appendix A) met Lawshe's method of validation.

After development, the survey was distributed to the standards-based grading teachers to understand their level of implementation of the standards-based grading practices they had been trained in. Teachers in the standards-based grading cohort answered the survey questions on their own and were asked to self-identify areas of strength and weakness. Of the 45 teachers in the standards-based grading cohort, 27 responded to the survey for a response rate of 61.4%.

Next, the survey results were de-identified by a member of the district data and

assessment, given to the researcher, and scored for implementation fidelity. Using the survey results, the researcher scored each question to determine which teachers in the standards-based cohort were high implementers or low implementers. Participants whose survey answers indicated they were following at least 90% of the prescribed grading strategies in implementing standards-based grading in their classrooms were flagged as high implementers (40%) in the standards-based grading participant cohort. Participants who were implementing less than 90% of the prescribed grading methods were flagged as low implementers (60%) in the same group. None of the teachers who responded to the survey were implementing less than 50% of the prescribed standards-based grading methods. Teachers in the standards-based grading cohort who did not respond to the survey were flagged as not having participated in the study, but their student data were still included in the standards-based cohort and listed as low implementers.

## **Phase Two**

The school district data and assessment team created a data file containing teacher information of those participating in the standards-based grading cohort. The file included school level taught (junior high or high school), years of teaching experience, subject taught, a descriptor of high implementer or low implementer, and whether they participated in the survey. Next, the data and assessment department found “like” teachers using the criteria of similar school demographics, grade level, subject matter, and teacher’s years of experience. Like teachers were found to create a control group to represent the traditional grading group. Data for 90 teachers were included in the study; 45 standards-based grading teachers and 45 traditional grading teachers.

Using the teacher's names, the data and assessment department compiled a list of students, their grade level, course subject, term grades in that teachers' respective class, SAGE scale score, SAGE growth percentage, and if the students were enrolled in that teacher's class for the entire school year. Data were deleted for students who switched teachers during the school year. We did not want data for students who participated in a standards-based grading classroom for half of the year and a traditional grading classroom for the other half of the year. Data were de-identified and each student was given a "dummy" student ID number. Grade and end-of-level SAGE test data were collected using an ex post facto design. The student data obtained were end-of-term classroom grades and end-of-level SAGE scales scores in math, science, and language arts. Data from the traditional grading group were collected with the intent of keeping it as similar to the standards-based grading cohort. The data and assessment department used school demographics, grade level, subject matter, and teacher's years of experience to match the traditional grading teachers to the standards-based grading teachers. All data collected were from the 2016-17 school year.

### **Data Analysis Procedures**

Because letter-based term grades are awarded to students, the researcher used a traditional GPA scale (Table 3) to convert letter-based term grades to a number. Each student received four term grades in every class, each school year. Each term letter grade was converted to a number, and the average of the four term grades was calculated yielding a GPA for each student which represents average graded performance during the

Table 3

*Grade Point Average (GPA) Scale*

Letter grade (A-F)	Assigned number (0-4)
A	4.00
A-	3.67
B+	3.33
B	3.00
B-	2.67
C+	2.33
C	2.00
C-	1.67
D+	1.33
D	1.00
D-	0.67
F	0.00

year. For example, if a student received a B+ first term, a C second term, an A third term, and a B- fourth term, the student's GPA for that course would be  $3.00 (3.33 + 2.00 + 4.00 + 2.67) / 4 = 3.00$ .

SAGE scaled scores in language arts and math are based on a vertical scale that changes with each grade level. SAGE scaled scores in science are based on a static scale that does not change with each grade level; however, students' SAGE scaled scores vary based on which science course a student is enrolled in. While each student is given a SAGE proficiency score of 1-4, scaled scores are more specific, and are therefore a more appropriate measure for this study than proficiency scores. Since SAGE scores are scaled differently across grade levels/content areas, and GPA scores are on a scale that differs widely from the scale of SAGE scores, a *z* score for SAGE scale scores was calculated for each student based on population means and standard deviations.

Student subjects were categorized into two distinct groups depending on whether they were enrolled in a standards-based classroom or a traditional classroom. The relationship between the two variables ( $z$  scores) were explored for each group using descriptive statistics and bivariate scatter plots with calculated Pearson correlation coefficients. Additionally, visual inspection of frequency histograms indicated whether the distribution of scores was approximately normal in each group of students. Two distinct independent samples  $t$  tests were conducted to determine whether a statistically significant difference in means exists between students enrolled in a standards-based classroom and students enrolled in a traditional classroom. The first  $t$  test determined whether GPA differed between the two groups. The second determined whether SAGE scaled score ( $z$  scores) differed between the two groups.

Since an acceptable predictive relationship between GPA and SAGE scores existed, a linear regression model was used to identify a predicted SAGE score based on each student's GPA. SAGE score residuals were calculated by subtracting the predicted SAGE score from the student's actual SAGE score. A third independent sample  $t$  test was used to determine whether SAGE score residuals differed between the two groups. These results allowed the researcher to determine whether there was a correlation between classroom assigned grades and end-of-level SAGE test scores.

Analyses were conducted separately for each of the three tested subject areas (math, science, and English) to determine whether standards-based grading had a stronger correlation with SAGE scores in tested subject areas. With the standards-based cohort acting as the independent variable and the student growth percentile acting as the

dependent variable, an independent samples  $t$  test was used to determine whether standards-based grading increased student learning as measured by the student growth percentage. Lastly, standards-based data were separated into groups based on years of teaching experience for educators in the standards-based grading cohort and the traditional grading group. Independent samples  $t$  tests were used to determine if a higher correlation of standards-based grades and end-of-level test scores varied based on years of teaching experience.

### **Conclusion**

In Chapter III, the methods of the study were presented. The researcher used the grade and end-of-level testing data to run several tests, including: Pearson correlation, frequency histograms,  $t$  test, and linear regression models. Assessing the correlation of grades and end-of-level test scores in standards-based classrooms and traditional classrooms adds to the limited literature that already exists. In addition, determining if the grading method used in classrooms affects student learning through the growth measure on end-of-level tests, is an area with limited existing research. Results from the proposed study helped determine if the traditional methods of classroom grading are effective or if a grading reform is needed.

## **CHAPTER IV**

### **RESULTS**

A large, suburban school district in the western U.S. recently began a standards-based grading experiment with secondary math, language arts, and science teachers. The objective of grading is to determine what a student has learned and what that student can do as a result of learning. Traditional grading methods often include data that are not related to what a student knows and can do. Standards-based grading is designed to enable educators to determine how well a student has learned specific content and skills related to specific learning standards.

The purpose of this study was to evaluate the effectiveness of determining student learning through the use of standards-based grading in secondary math, language arts, and science classrooms and the correlation those grades have with end-of-level test scores. Determining student learning is a difficult task that educators have struggled to do effectively for many years. There are many variables that factor into student learning that are beyond a researcher's control. While various factors are difficult to control, this study focuses on the two main indicators used in public schools to determine student learning: term grades and end-of-level test scores. By analyzing these two data points, results will add to the ongoing discussion of determining student learning. This section presents the results of the analysis. Results are ordered by research question with a summary of results at the end of each question's section.

### Correlation Between Grades and End-of-Level SAGE Test Scores

Several tests were completed to answer this research question. First, descriptive statistics (Table 4) were calculated for both the standards-based cohort and the traditional group. Before testing a regression, a visual inspection of bivariate scatter plots with calculated Pearson correlation coefficient for the standards-based cohort ( $r = .492$ ) and the traditional group ( $r = .530$ ) were run to determine if the variables have a linear relationship. Results indicate student GPA and SAGE  $z$  scores have a linear relationship for both groups.

Frequency histograms were created and visually analyzed to determine whether the distribution of scores is approximately normal. SAGE  $z$  score frequency histograms of the standards-based cohort and the traditional group were created and are approximately normal. The frequency histograms representing the GPA for both groups are not normal; however, distribution for both groups are visually similar. Though the histograms are skewed, Pearson's correlation is a dependable estimator of the population correlation even when bivariate normality is not represented.

Table 4

#### *Standards-Based and Traditional Grading Groups Descriptive Statistics for GPA and SAGE $z$ Scores*

Group	Indicator	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>	Variance
Standards-based cohort	GPA	4,107	.000	4.000	3.165	.966	.933
	SAGE $z$ score	4,107	-3.526	2.909	.147	.999	.997
Traditional group	GPA	4,022	.000	4.000	3.008	1.044	1.090
	SAGE $z$ score	4,022	-4.234	2.647	-.150	.979	.959



Independent samples *t* tests for each group were conducted to determine if a statistically significant difference exists between the groups' GPA and SAGE scaled *z* scores. Results of the first independent samples *t* test shows a significant difference of the mean GPA scores between the standards-based grading cohort ( $M = 3.165$ ,  $SD = .966$ ,  $n = 4107$ ) and traditional grading group ( $M = 3.008$ ,  $SD = 1.044$ ,  $n = 4022$ );  $t(8048.718) = -7.015$ ,  $p < .005$ . Students in the standards-based grading cohort scored a mean difference of .157 GPA higher than their peers in the traditional grading group; meaning students in the standards based cohort scored higher average grades in tested subjects than students from the traditional grading group. Though the results are significant, the difference between the two groups is nominal. Numerically, students in the standards-based cohort receive a half grade higher than their peers in the traditional grading group. In most 100-point grading scales, this would be the difference of about 2.5% percentage points. The half grade represents half of the difference between a B- and a B, or a B+ and an A-.

Results of the second independent samples *t* test shows a significant difference of the mean SAGE scaled *z* scores between the standards-based grading cohort ( $M = .147$ ,  $SD = .999$ ,  $n = 4107$ ) and traditional grading group ( $M = -.150$ ,  $SD = .979$ ,  $n = 4022$ );  $t(8126.990) = -13.507$ ,  $p < .005$ . On average students in the standards-based grading cohort scored higher on end-of-level SAGE tests by .293 standard deviations – a significant and meaningful difference – than their peers in the traditional grading group. In addition, error bars (Appendix B) were created for both the GPA data and the SAGE *z* score data showing a visual representation for the *t* tests. Both charts reiterate at least a 95% confidence level and show a significant difference for each data point.

Correlation tests were used to determine the relationship between student grades and SAGE test scores. Participants in the standards-based grading cohort ( $r = .492$ ) and the traditional grading group ( $r = .531$ ) both show a meaningful relationship at a significant level ( $p < .005$ ). A linear regression model was used to determine if there is a correlation between classroom assigned grades and end-of-level SAGE test scores. Descriptive statistics for the regression analysis are shown in Table 5. In addition, a residual standardized histogram (Appendix C) was completed to visually analyze if the distribution of scores is approximately normal. The R Square score for all cases was .266, meaning 73.4% of students' SAGE scores can be attributed to something other than the end-of-term grades in that subject. The regression ANOVA results are in Table 6. Results from coefficients show that if a student can increase their grade by one GPA point, from a C to a B or from a B to an A, their SAGE scores will increase by half a standard deviation ( $\hat{y} = -1.578 + .511[\text{GPA}]$ ).

Table 5

*Regression Model: Descriptive Statistics for GPA and SAGE z Scores*

Indicator	<i>N</i>	Mean	<i>SD</i>
GPA	8,129	3.087	1.001
SAGE <i>z</i> score	8,129	.000	1.000

Table 6

*Regression ANOVA Model: GPA and SAGE z Scores*

Indicator	Sum of squares	<i>df</i>	Mean square	<i>F</i>	Sig
Regression	2159.877	1	2159.877	2940.686	.000
Residual	5969.123	8127	.734		
Total	8129.000	8128			

Results of the unstandardized regression residuals independent samples  $t$  test shows a significant difference of the mean SAGE scaled  $z$  scores between the standards-based grading cohort ( $M = .107$ ,  $SD = .869$ ,  $n = 4107$ ) and traditional grading group ( $M = -.109$ ,  $SD = .830$ ,  $n = 4022$ );  $t(8121.826) = -11,473$ ,  $p < .005$ . Based on this data, neither group provides a more accurate correlation to predict SAGE scores with course grades. The GPA in the standards-based grading cohort is under predicting the SAGE score by .107 and the GPA in the traditional grading group is over predicting by .109 (Appendix D)

In addition to comparing data for the standards-based grading cohort and the traditional grading group, the researcher looked at the difference between participants in the standards-based cohort who self-identified themselves through a survey as a high or low implementer of standards-based grading. Teachers in the standards-based cohort were asked to complete the Standards-Based Grading in the Classroom survey (Appendix A) and evaluate themselves in specific standards-based grading criteria. The data from the surveys were coded and respondents were separated into two groups: low implementers and high implementers. Data from these groups were separated and a  $t$  test was run to determine statistics for each group. Table 7 provides the statistics for each group. Results of the independent samples  $t$  test looking at the unstandardized residual of the GPA for low implementers of the standards-based grading cohort and high implementers of the standards-based grading cohort show a significant difference of the mean GPA scores;  $t(3791.740) = -4.611$ ,  $p < .005$ . The mean difference shows the GPA for high implementers to be .139, or close to one-third of the difference in a grade

change. Additionally, results of the independent samples  $t$  test looking at the unstandardized residual of the SAGE  $z$  scores for low implementers of the standards-based grading cohort and high implementers of the standards-based grading cohort show a significant difference of the mean SGP scores;  $t(4105) = 5.740, p < .005$ . Interestingly, end-of-level SAGE results were higher for the self-identified group of low implementing teachers. The mean difference between the two groups was .181.

Summarizing question 1, there is a correlation between standards-based grades and end-of-level SAGE scores; however, the correlation is not more accurate for the standards-based grading cohort than it was for the traditional grading group. Grades and SAGE test scores in the standards-based cohort were higher than the grades and SAGE test scores in the traditional group, but neither group had data to support a more predictive relationship. When the standards-based grading data were separated by self-identified low and high implementers, results show a higher GPA with the high implementing group, but a lower end-of-level SAGE  $z$  score. Based on this data, the researcher cannot conclude that standards-based grading is more correlated with end-of-level SAGE test scores than with traditional grading methods. However, data do show

Table 7

*Descriptive Statistics for GPA and SAGE  $z$  Scores Based on Level of Implementation (Self-Identified)*

Indicator	Fidelity (self-identified)	$N$	Mean	$SD$
GPA	Low	2,463	3.109	1.007
	High	1,644	3.248	.894
SAGE $z$ score	Low	2,463	.219	1.003
	High	1,644	.038	.982

students to have higher grades and higher end-of-level SAGE test scores when participating in a standards-based grading environment. The value of this may indicate the use of standards-based grades may indicate improved student learning. This will be further discussed in Chapter V.

### Effect on Student Learning

Student growth percentiles (SGP) are a way of quantifying the academic growth a student has made compared to their peers across the state of Utah as a result of one year of instruction. SGP are calculated using the previous year's end-of-level SAGE test score in each content area. Students are arranged into an academic peer group of students who scored the same as them on the SAGE test. SAGE scores for the next year are then compared with students' scores from the same academic peer group. A score of 40 SGP means a student made more growth than 40% of the students from their academic peer group. A score of 50 SGP represents average growth. Descriptive statistics for both groups show above average growth (Table 8). Results of the independent samples *t* test comparing the standards-based grading cohort and the traditional grading group show a significant difference of the mean SGP scores;  $t(7594) = -6.669, p < .005$ . Numerically,

Table 8

#### *Descriptive Statistics of Student Growth Percentiles for Both Groups*

Indicator	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>	Variance
Standards-based cohort	3,790	1	99	55.590	28.621	819.148
Traditional group	3,806	1	99	51.190	28.562	815.792

students in the standards-based grading cohort had 4.4% higher SGP than students in the traditional grading group. The difference in SGP are significant and indicate a meaningful effect size between student groups. An error bar (Appendix E) was run to test for confidency levels.

An additional  $t$  test was used to identify any differences between the self-identified low and high standards-based grading implementers. The descriptive statistics for each group can be found in Table 9. Results from the  $t$  test did not result in being statistically significant as the mean difference was .21:  $t(3788) = .183, p > .05$ .

In sum, the results for Question 2 are significant in that the standards-based grading cohort had a significantly higher SGP than their peers across the state in the traditional grading group. While both the standards-based and traditional based grading groups showed a higher than average 50% SGP, the standards-based cohort had an SGP of 4.4% higher than the traditional grading group. The SGP for teachers who participated in the standards-based grading cohort and identified themselves as low or high standards-based grading implementers only differed by 0.210.

### **Difference by Core Subject**

Participants from the standards-based grading cohort included teachers from the

Table 9

#### *Descriptive Statistics for SGP Based on Level of Implementation (Self-Identified)*

Indicator	Fidelity (self-identified)	$N$	Mean	$SD$
SGP	Low	2,262	55.66	28.806
	High	1,528	55.45	28.354

language arts, math, and science content areas. Often, teachers implementing standards-based grading practices from these content areas believe their content may or may not work with different grading practices. Independent samples  $t$  test regression residuals and error bars were created for each content area looking at the correlation of content GPA and SAGE  $z$  score residuals.

Statistics results for the language arts group are found in Table 10. Results of the independent samples  $t$  test looking at the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(3472.076) = -7.670, p < .005$ . An error bar chart was created for the language arts group (Appendix F). The SAGE scores in the standards-based grading cohort are underestimated by .115 and the SAGE scores in the traditional grading group are overestimated by .105.

Statistic results for the math group are found in Table 10. Results of the independent samples  $t$  test looking at the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(2614.697) = -15.999, p < .005$ . An error bar chart was created for the math group. The GPA in the standards-based

Table 10

*Content Area Statistics Regression Residuals-SAGE  $z$  Score with GPA*

Subject	Indicator	$N$	Mean	$SD$	Std. error mean
Language arts	Standards-based cohort	1,737	.115	.873	.021
	Traditional group	1,751	-.105	.826	.020
Math	Standards-based cohort	1,230	.377	.752	.021
	Traditional group	1,397	-.109	.802	.022
Science	Standards-based cohort	1,140	-.197	.885	.026
	Traditional group	874	-.118	.882	.026

grading cohort is underestimating the SAGE score by .377 and the GPA in the traditional grading group is overestimating the SAGE score by .109 (Appendix F). The data indicate that using standards-based grading in math classes produces grades that are less correlated with end-of-level SAGE test scores than using traditional grading methods.

Statistics results for the science groups are found in Table 10. Results of the independent samples  $t$  test looking at the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(2012) = 1.978, p < .05$ . The GPA in the standards-based grading cohort is overestimated by .197 and the GPA in the traditional grading group is overestimated by .118 (Appendix F).

Summarizing question 3, results from the sample  $t$  test residuals and error bar charts for each subject show language arts as the content area that provided the closest correlation for GPA and SAGE test scores for the standards-based cohort. Math was the lowest correlated. Interestingly, science was the only content area with an overestimated confidency level in the standards-based cohort, while all three content areas overestimated the confidency levels in the traditional grading group.

### **Difference by Years of Teaching Experience**

Years of teaching experience can have an impact on many areas of education. For this study, all the teachers in the standards-based grading cohort have been implementing standards-based grading for the same period of time. However, there is a wide range in the years of teaching experience for both grading groups.

Group statistics results for educators with 0-5 years teaching experience are found



in Table 11. Results of the independent samples  $t$  test looking at the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(3564.947) = -3.500, p < .005$ . The GPA in the standards-based grading cohort is overestimated by .030 and the GPA in the traditional grading group is overestimated by .133 (Appendix G).

Group statistics results for educators with 6-10 years teaching experience are found in Table 11. Results of the independent samples  $t$  test observing the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(1832.821) = -13.131, p < .005$ . The SAGE scores in the standards-based grading cohort are underestimated by .266 and the SAGE scores in the traditional grading group are overestimated by .227 (Appendix G).

Group statistics results for educators with 11-15 years of teaching experience are found in Table 11. Results of the independent samples  $t$  test looking at the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP

Table 11

*Teaching Experience Group Statistics Regression Residuals-SAGE  $z$  Score with GPA*

Experience	Indicator	$N$	Mean	$SD$	Std. error mean
0-5 Years	Standards-based cohort	1,938	-.030	.895	.020
	Traditional group	1,677	-.133	.870	.021
6-10 Years	Standards-based cohort	1,108	.266	.802	.024
	Traditional group	884	-.227	.858	.029
11-15 Years	Standards-based cohort	721	.124	.886	.033
	Traditional group	573	-.074	.736	.031
16+ Years	Standards-based cohort	340	.334	.755	.041
	Traditional group	888	.030	.761	.026

scores;  $t(1289.472) = -4.387, p < .005$ . The SAGE scores in the standards-based grading cohort are underestimated by .124 and the SAGE scores in the traditional grading group are overestimated by .074 (Appendix G).

Group statistics results for educators with 16+ years of teaching experience are found in Table 11. Results of the independent samples  $t$  test analyzing the unstandardized residual of SAGE  $z$  scores with GPA show a significant difference of the mean SGP scores;  $t(1226) = -6.270, p < .005$ . The GPA in the standards-based grading cohort is underestimated by .334 and the GPA in the traditional grading group is overestimated by .030 (Appendix G).

Summarizing question 4, the correlation between SAGE  $z$  scores and class GPAs differed with each group of educators' teaching experience. The group of teachers with 0-5 years teaching experience showed the standards-based grading cohort to have a higher correlation and better predictor of SAGE  $z$  scores and GPA. The group of teachers with 6-10 years of experience show very little difference in correlation between the two grading groups. The standards-based cohort underestimated GPA, while the traditional grading group overestimated the GPA. The groups of teachers who have 11-15 years' experience and 16+ years' experience showed that the traditional grading group has SAGE  $z$  scores and GPAs that are more correlated than the same data from the standards-based grading cohort.

## Conclusion

This chapter included the results from the tests run using the SPSS statistical

software. Results for each question were presented individually. The results for the standards-based grading cohort appear to produce higher GPAs, end-of-level SAGE scores, and more growth on the SAGE test. Correlation between grades and end-of-level test scores varied based on the subject and teachers' years of experience. The final chapter will include a discussion of the study's finding, future research recommendations, and limitations of this study.

## **CHAPTER V**

### **DISCUSSION**

During the 5 five years, I have had the opportunity to be an administrator at three different secondary schools. While each school has a unique culture and different challenges, one thing is always the same—teachers demonstrated inconsistent grading practices. I have had innumerable meetings with parents, students, and teachers regarding grading practices. After a long meeting with a teacher and parents regarding a student's grade, I was driving home and realized something needs to change.

Let me start at my beginning. I was a very good student. I was never the smartest student in my classes, but I knew how to play the “school game” of being respectful and getting my work in on time. In ninth grade, I missed a week of school due to an illness. During my week at home, I got very behind in my math class. In fact, I got so behind that I was never able to catch up. Too scared to ask for help, I turned to the next best thing, cheating. My best friend completed the homework every night and allowed me to copy his homework every morning before math class. I received close to 100% on every assignment and failed every test. At the end of the school year, I finished with a B average in my math class. Had my teacher been watching, she would have realized I had no idea what I was doing. The scores I earned on assessments were abysmal, yet there were no interventions provided.

Skipping forward several years to the present, I have completed two bachelor's degrees, a master's degree, an administrative endorsement, and coursework for a Ph.D. I have been a classroom teacher, department chair, coach, and a school administrator.

Sadly, I have never been trained on how to assess and grade effectively. The facts are not a knock on the universities I attended, my professors, or the school leaders I have worked with. I have been very lucky to learn from some of the greatest people I have ever met. This observation is something I think most educators can relate to. Most teachers are not being trained to implement effective grading practices before entering the teaching field, nor are they being trained in their schools. A couple times a year, a new teacher will come into my office and ask if he or she can give a failing grade to a student because of any number of things, including cheating, being disrespectful to a substitute teacher, and leaving class early. While all of these things are concerning, they do not reflect what a student knows and should not be put into a student's grade.

Most teachers were good students who tried their best, liked school, and decided to spend a lifetime working in a school setting. Because of their success in school, these teachers like the system and do not see reasons to change it. Effective grading practice is one area rarely considered when discussing change. Teachers grade their students the same way they were graded in school. It becomes a cycle of mediocrity (Guskey, 2015). It is time for the cycle to be changed. I would like to go through three situations I have encountered over the last five years to show that a grading reform is a hot topic in education.

Early in my career as a school administrator, there were several teachers who gave failing grades to a high percentage of students. When one teacher was asked why so many students were failing her class, she replied, "My class is harder than most because I am a great teacher and do not give grades away for free." As our discussion continued,

we talked about the traits of a great teacher. We discussed great teachers this teacher had in her life. We settled on the fact that great teachers inspire students to do their best-while motivating and helping all levels of learners to reach their potential. They do not just teach; great teachers ensure students learn. The next question a colleague asked the teacher was, “If you are such a great teacher, how come so many students are not learning?” The teacher replied, “Students are learning in my class”, to which my colleague replied, “Then why do their grades not reflect what they have learned?”

The second situation that has had a lasting impact on me in regard to grading concerns a group of teachers who heard the school district was potentially going to mandate a change to standards-based grading, and they wanted to get ahead of the curve. These teachers did some research online and began implementing bits and pieces of standards-based grading. Classroom disclosures were updated to reflect the grading change that homework was strictly for practice and tests and quizzes would account for 100% of a student’s grade. Unfortunately for students, they were not allowed to retake assessments. Whatever grade the student earned on the test, was the grade received at the end of each term. Students did not have a method to improve their grades through retesting. Each school administrator had multiple meetings with parents to discuss concerns about how unfair the grades were. Because of the parents’ and students’ frustration, this group of teachers went back to their old way of grading.

The last situation involves a group of teachers who realized on their own that student grades did not reflect what students had learned. They came to the school administration and asked for money to attend a professional development conference on

effective grading. Upon returning, the group of teachers began formulating a plan to effectively implement their new grading practices. Before the school year started, parents were notified about the grading policies and were also provided an information sheet. As the school year progressed, some things did not work as expected so changes were discussed and made quickly and efficiently. Parents and students were kept in the loop about any changes that affected them. While some parents initially did not like the grading practices, their mood changed as they saw how motivated their children were to learn and re-learn concepts so they could retest. Students began to focus on relearning standards, instead of getting extra credit to raise their grade.

I bring these three situations up because they show the diverse thoughts teachers in a school building have about grading. Schools will have teachers who think it is their life mission to make students suffer, teachers who half-heartedly implement changes, only to see them fail and go back to what is comfortable, or teachers who will look for ways to make positive changes by being proactive and including stakeholders in the changes. The last group of teachers realizes that change will not be easy, and changes will need to be made along the way, but it is what is best for students.

To effectively implement a grading change, the school culture must be ready for it. That is not to say, nothing can be changed until every single person is on board, but key people need to be willing to implement the changes. School administrators need to ensure the right people are in leadership positions. No matter how good an idea or strategy might be, culture trumps strategy.

My hope is that this research will help educational leaders make informed

decisions in regard to implementing grading changes in their schools. Before making a change, leaders need to ensure they are knowledgeable on the subject themselves by researching beforehand. Having knowledge will help ease teachers' concerns as they work through the process of change. Second, school leaders need to ensure their school culture is ready for the change. Educators are all too familiar with change and many teachers have grown weary of policy changes. Leaders should ensure leadership teams are ready and willing to change so they can be a voice of reason in the faculty room. Third, school leaders need to provide professional development to teachers. Most educators have not been trained on grading practices and it is unreasonable to think they will figure it out on their own. Resources must be allocated to help ensure teachers are knowledgeable about the subject and know where to go when questions arise. Lastly, school leaders must educate and inform stakeholders. Parents are familiar with traditional grading practices and are comfortable with them. A change in grading practice is no small ordeal and information must be disseminated. Educational reforms are not a small change and a grading reform is no different. They are a change that requires years of hard work. Education leaders need to be prepared beforehand if the reform is going to be effective.

### **Overview**

The purpose of this study was to examine the effects of a standards-based grading experiment on student learning through the use of student grades and end-of-level SAGE test scores. During the 2015-16 school year, a large, suburban school district in the



western U.S. began a 2-year standards-based grading experiment with secondary language arts, math, and science teachers. Grade and end-of-level testing data from teachers in the standards-based grading cohort were analyzed and compared with data from a group of teachers using traditional grading methods. In this chapter, we will review the issues with traditional grading practices and the research questions, discuss the research results, and provide information and tips for educators to use as they seek to improve grading practices in their own school. This research will prove to be for nothing if changes in current grading practices do not occur. Students deserve to be graded accurately and fairly.

### **Review of the Research and Research Questions**

For over a century, the traditional form of grading has remained unchanged. Students are being assessed on items unrelated to learning, such as behavior, attendance, work completion, and time spent on homework. While these are important skills and traits to develop, they do not inform the student or parent what the student knows and what they can do.

Education has recently increased its focus on having common standards. These standards are meant to ensure students across the country are gaining the same knowledge and skills. Unfortunately, a common form of grading does not exist and student grades mean different things, depending on the grading philosophy of the teacher teaching the course. It is important educators come to a consensus about using effective ways to determine a student's grade. Students across the world are competing against

each other for awards, college enrollment, and scholarships. They may be learning common standards and objectives, but they are not being assessed in a communal way. The standards-based grading cohort was formed to help identify if standards-based grading is an effective way to determine student learning.

To determine if standards-based grading is an effective form of grading, a survey was created for teachers to take in the standards-based grading cohort to help determine how closely teachers were following standards-based grading procedures. Next, we used the data from the survey to remove teachers from the study who were not implementing prescribed grading techniques. Then, the researchers identified a group of teachers who did not participate in the standards-based grading cohort but taught in the same subject, grades, and similar school demographics compared to the standards-based grading teachers. Data were obtained and de-identified through the school district database, and the data were analyzed and compared to answer the following research questions.

1. Is there a positive correlation between standards-based grading, term grades, and end-of-level SAGE test scores?
2. To what extent does the standards-based grading environment increase or decrease student learning, in comparison with traditional grading, as measured by the student growth percentage on end-of-level SAGE test scores?
3. Is there a difference in the correlation of standards-based grades and end-of-level SAGE test scores between math, language arts, and science?
4. Does the correlation between standards-based grades and end-of-level test scores vary based on teachers' years of teaching experience?

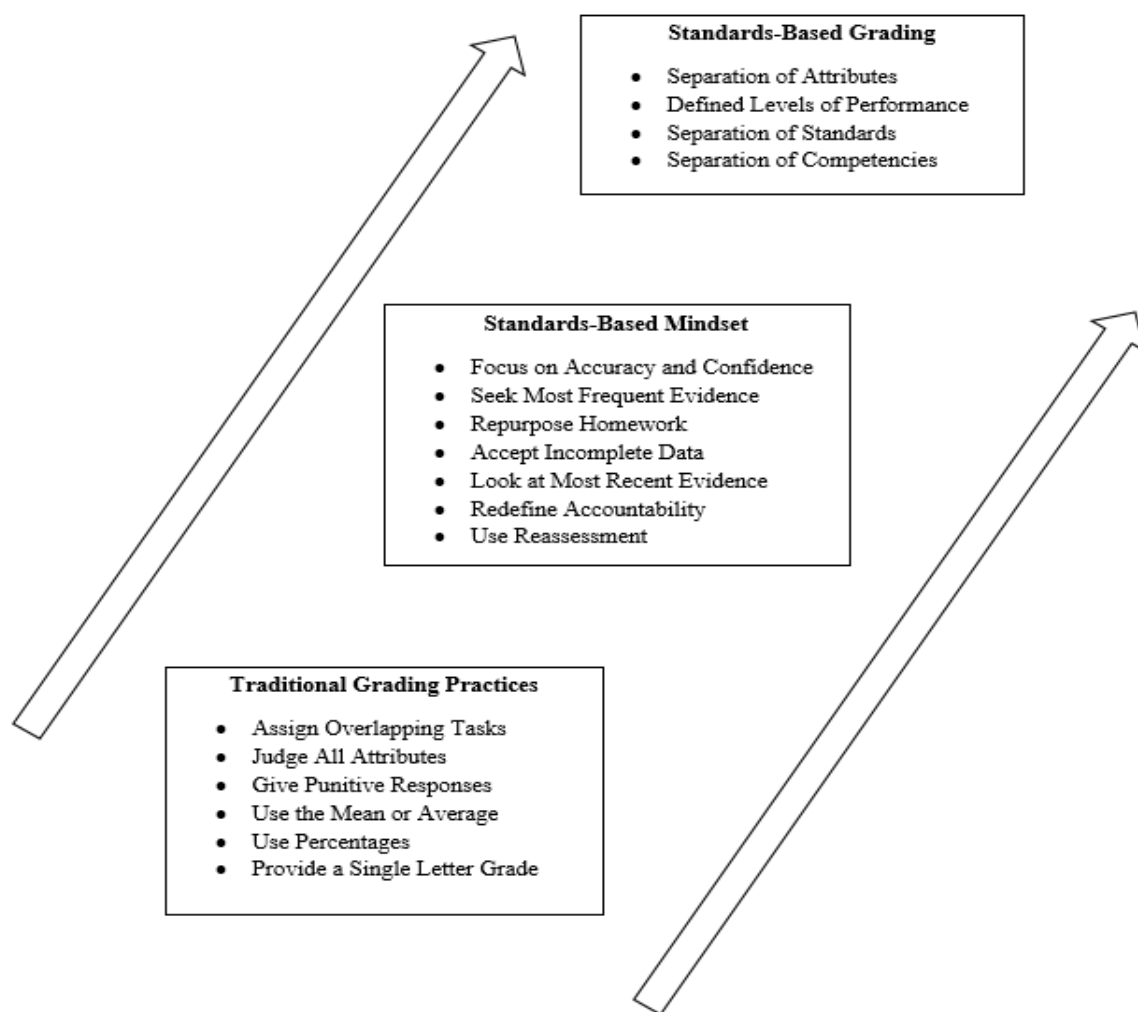
### **Fidelity of Implementation**

Research fidelity is the level to which execution of an intervention follows the

program model originally established (Dusenbury, Brannigan, Falco, & Hansen, 2003). Using this definition, the researcher examined, through a self-reporting survey, the extent to which teachers in the standards-based grading cohort were implementing the strategies and practices they had learned through the professional learning experiences they engaged in. Specifically, assistance and guidance were given to the cohort through a book study and collaboration meetings and presentations.

Survey results showed teachers who responded to the survey (61.4%) in the standards-based grading cohort indicated 40% of the teachers were high implementers, applying at least 90% of the prescribed standards-based grading methods. The other 60% of teachers who responded to the survey, showed they were implementing less than 90% of the standards-based grading methods. These teachers were classified as low implementers. None of the survey respondents indicated they were implementing less than 50% of the prescribed standards-based grading methods. Overall, these results show teachers in the standards-based grading cohort executed the intervention with fidelity as prescribed by their district leaders.

A change in grading practices requires professional development and time for a lasting change to occur. Schimmer (2016) explained the gap from traditional grading practices to standards-based grading needs to be filled with a standards-based mindset (Figure 1). This mindset allows educators to view grading from a standards-based assessment. The standards-based grading mindset is important because educators cannot wait for a perfect grading system and tools before making changes to grading practices. Standards-based grading mindset is process that can be started one step at a time.



*Figure 1.* Standards-based grading mindset. Adapted from (Schimmer, 2016).

## Discussion of Results

The study consisted of examining student grades and end-of-level SAGE data from the standards-based grading cohort, in comparison with student data whose teachers did not participate in the grading cohort. Overall, students who participated in the standards-based grading cohort earned higher grades and scored a higher proficiency and higher growth on the end-of-level SAGE tests. Major limitations include time and the

inability to determine if students took the end-of-level SAGE test seriously. In chapter 5, we discuss the results of the study and the implications the study has on grading in education. The discussion items are separated by individual research questions. The limitations, future areas of research study, and conclusion follow the discussion for each research question.

### **Correlation of Standards-Based Grading Term Grades and SAGE Scores**

To answer whether standards-based grades have a positive correlation with end-of-level SAGE test scores, we also had to address and identify the relationship between traditional grades and end-of-level SAGE test scores. After analysis of the data, both grading groups' grades proved to be statistically correlated with end-of-level tests, however, some other interesting results were found.

First, comparing the standards-based grading cohort with the traditional grading group found that students who participated in the standards-based cohort earned higher overall grades in their respective math, language arts, and science classes. In fact, the grades were half the difference between a letter grade; or half the difference between a C- and a C, or a C and a C+. In a traditional 100% grading scale, this difference is equal to a 2.5% score increase in comparison to students whose teachers were included in the traditional grading group. While the 2.5% percent difference may seem small on a 100-point scale, it can be the difference between a grade change, which increases the students' overall GPA, allowing them to increase their chances to be eligible for honor rolls, college enrollment, and scholarships.

Having an increase in grades is a big deal in education. Much of an educational leader's job is meeting with parents who have concerns about their child's grade. When parents can walk away from a meeting feeling like their child is being graded fairly and has opportunities to improve upon their learning, parents are typically happy. As demonstrated in this study, as teachers assess on true learning standards and provide multiple chances for students to demonstrate learning, grades will increase.

In addition, students in the standards-based cohort proved to have higher end-of-level SAGE scores in each respective subject. The difference in SAGE scores was the standards-based grading cohort was  $\frac{1}{3}$  of a standard deviation better than the traditional grading group. As with the difference between classroom grades, the difference in SAGE scores was also significant in that students in the standards-based grading cohort scored 11.5% percent higher on the SAGE test than their peers in the traditional graded classes. Scoring higher on the SAGE test does not result in a higher grade in the class, but it does show students are more proficient on the learning standards and are more prepared for the upcoming school year.

Achieving higher scores on high stakes tests is important for students as they seek to enroll in college and earn scholarships. Colleges typically look at two things; grades on a student's transcript and results from a high stakes tests like the ACT or SAT. These two learning indicators are the two key components colleges use to determine academic scholarship opportunities.

In sum, there was a positive correlation between grades and end-of-level tests from both the standards-based and traditional grading groups. With that in mind, student

grades and SAGE test scores were higher in classrooms that used standards-based grading as the grading measure. Earning higher grades and performing better on the SAGE test is a sound indicator to demonstrate students learned more in the standards-based grading classes. The increased performance on grades and end-of-level SAGE for the standards-based grading cohort could be attributed to the focus standards-based grading provides on ensuring students are reaching for proficiency in each standard. If students do not show proficiency, they are required to be re-taught and reassessed until proficiency is shown. Having the ability to retake tests and keep the highest score will naturally increase grades. End-of-level SAGE test scores are aligned with learning standards, so ensuring students are proficient in those standards throughout the school year will help students focus on learning. Earning higher grades and showing proficiency with learning standards is important for students as they prepare for college admission and scholarships.

### **Influence of Grading Practices on Student Growth**

To measure the difference in student growth and answer research question #2, growth scores from end-of-level SAGE tests were used. The student growth percentile measures the amount of academic growth a student has made as a result of one year of instruction when compared to equally performing peers. The growth calculation identifies an academic peer group of students who performed at the same level in the prior year, and produces a growth score depending on how the peer group performed on SAGE in the current year. Student growth measures learning from year to year in comparison to

similar performing students.

While both groups showed an above-average growth, the standards-based grading cohort demonstrated 5% more growth than the traditional grading group. Student growth is an important statistic because it demonstrates the learning increase over the course of a school year. As previously mentioned, teachers cannot affect how a student enters a class, but they do have an impact on how much a student grows intellectually throughout the school year (Newton & Winches, 2013). Embracing the idea that all students can learn and make progress is an essential mindset for teachers to have. In a typical classroom, a teacher is tasked with instructing students who have different capabilities and learning styles. Teachers must focus on reaching all levels of learners and ensure learning is taking place all the time. Standards-based grading in the classroom promotes student learning and growth in particular areas, according to the data from this research. As assessments are given and data are obtained, teachers have the responsibility of providing feedback to students. If a student has not achieved proficiency on a standard, re-teaching and reassessing are essential next steps in the learning process.

Student growth from year to year is incredibly important. In fact, a primary indicator of effective teachers and school administrators is the learning growth students have during the school year (USDE, 2009). As students enter a classroom, teachers are tasked with instructing students who are on different learning and skill levels. Some students in the classroom come into the school year ahead of the grade-level requirements, while other students enter the school year multiple grade levels behind. While a teacher cannot affect how students come into their classes, a teacher can affect



how much progression and growth a student makes during the year (Newton & Winches, 2013). Student proficiency levels are often related to a variety of factors, including socio-economic status, parent support, and home life (Jensen, 2009).

The increased growth demonstrated in the standards-based cohort can be attributed to a variety of factors. A common practice in standards-based grading is the focus on ensuring proficiency. If a student does not show proficiency on an assessment, the student should be retaught and reassessed until the student can show proficiency to the standard. Unfortunately, re-teaching and reassessing are not essential steps in traditional grading practices. Failure to provide additional opportunities to relearn and reassess learning standards creates a problem for a curriculum that builds or scaffolds upon itself and requires learners to understand standard one before they can understand standard two (Iamarino, 2014). The attention and importance of showing proficiency for each standard in the standards-based grading cohort is an essential component of increasing student growth scores. Increasing student learning growth from year to year benefits all levels of learners. Students who come into a school year ahead of the grade level can continue to exceed grade-level expectations and students who begin the year below grade level can make progress toward closing the learning gap.

If teachers are able to help close learning gaps and increase learning each school year, graduation rates will continue to rise. In my experience, most students drop out of school due to the fact they are not being successful in school. Not feeling successful leads to students not attending school and they get left behind. If more students are able to learn each year, our economy will benefit by having a more educated society.

In sum, the standards-based grading cohort showed 5% more growth than the traditional grading group on the end-of-level SAGE test. The above-average growth demonstrated by the standards-based grading cohort could be a correlation with focus on earning proficiency in each standard that standards-based grading provides. In standards-based grading, ensuring students are learning and progressing in each standard is an essential component. By providing a learning environment with high expectations of ensuring students are proficient in learning standards, students are given the time and resources needed to make learning growth.

### **Correlation of Grading and SAGE Scores by Content Area**

While grading can vary from teacher to teacher, grading can also vary by curriculum content areas. Parents are often concerned their child can earn an A with one teacher and an F with a different teacher, simply because one teacher is a harder grader than the other. As it relates to grading, math is often considered to be clearer in assessing learning because students either acquire the correct answer or they do not. On the other hand, grading a writing paper can be very subjective based on the viewpoint of the person grading the paper. Question #3 focused on the correlation of content specific grades and end-of-level SAGE test scores as they relate to different curriculum areas: language arts, math, and science.

Correlation between grades and end-of-level test scores varied by subject and by grading group. Based on the results of this study, it is impossible to determine why one subject would be more correlated than another. It is possible some teachers did not follow

prescribed grading practices leading to grades that were higher or lower than what the students' learning actually demonstrated.

The first curriculum content area we looked at was the language arts group. Results from the data analysis showed that neither the standards-based grading cohort nor the traditional grading group showed more predictability between the grade given and the end-of-level SAGE test score (Appendix F). The standards-based grading cohort assigned student grades that would have predicted lower SAGE scores than what students actually demonstrated. However, the SAGE scores ended up being higher than the data from the predictive test would have projected based on the grade given in the class. The traditional grading group had the opposite effect. Grades assigned in the traditional grading group should have yielded higher SAGE test scores than students actually demonstrated on the test.

In curriculum area of math, the standards-based cohort SAGE scores were higher than what the students' grades predicted. The math SAGE scores from the traditional grading group were almost equal in predictability to the grades and SAGE test scores from the language arts group (Appendix F). Along with the language arts traditional grading group, the math SAGE scores were lower than the students' grades would have predicted. In the math group, the traditional grading group showed student grades and the SAGE test scores were more correlated than their standards-based grading peers.

Science was the only content area where both grading groups assigned grades that predicted students would receive higher SAGE scores than they actually did. Both groups assigned student grades that over predicted the SAGE scores students actually scored on

the test (Appendix F). As with the math group, the traditional grading groups' grades were more correlated with SAGE scores than the standards-based grading cohort.

Interestingly, the traditional grading group's data were relatively consistent from subject to subject, while the standards-based grading cohort varied considerably. In all three curriculum content areas the traditional grading group assigned grades that predicted a higher SAGE test score than what students actually received. The regression residual data for all three content areas were very close.

On the other hand, data from the standards-based grading cohort were very different from subject to subject. Language arts and math assigned grades that under predicted the SAGE test scores, while science assigned grades that over predicted the SAGE scores. The standards-based grading math cohort was an outlier compared to the other data, in that it's predictability of grades and SAGE scores differed the most.

Without completing additional research, it is impossible to determine exactly why the standards-based grading cohort differed so much from subject to subject, but traditional grading practices have been around for a long time. In most cases, teachers are more familiar with traditional grading practices because it is how they were graded when they attended school and how they have been grading for most of their careers (Frary et al., 1993; Guskey & Bailey, 2001). The standards-based grading cohort has only been implementing standards-based grading procedures for 2 years. There is still a great deal to learn and changes to be made. Additional time, resources, and professional development are needed to determine if standards-based grading can prove to be more predictive of student performance on end-of-level tests than traditional grading in the

future.

Another interesting outcome to our third research question is that teachers who use traditional grading consistently assign grades that overestimate student performance on end-of-level tests, meaning the grades they assign are inflated in comparison to test scores. Assigning grades that overestimate student test performance can possibly be attributed to the availability of extra points students can earn in a traditional classroom setting. In a study by Randall and Engelhard (2010), survey results showed student grades increased as behavior improved, regardless of academic ability. All these non-academic grading factors add to the concerns raised in the past about grading, including the misinterpretations of report cards, non-academic factors that influence grades, and the unreliability of percentage systems (Brookhart, 2004). Teachers' grading practices showed that nonachievement factors were regularly calculated into students' grades (Duncan & Noonan, 2007). While most teachers believe grades should reflect learning and achievement, many of those same teachers included points for good behavior, extra credit, and studying as part of a student's grade (Feldman et al., 1998).

In sum, the correlation results of assigned grades and end-of-level SAGE tests scores varied by curriculum content area. The one commonality obtained from this question is teachers from the traditional grading group assigned term grades that were inflated when compared to student end-of-level test scores. The grading practices teachers use may jeopardize the reliability of grades and therefore weaken the link between grades and academic achievement. That link weakens as teachers use grading practices unrelated to student learning (Welsh et al., 2013). If traditional grading methods

consistently show grades are inflated and do not accurately reflect student learning, educators must continue to look for strategies to improve. Standards-based grading might be the answer, but more time, professional development, and research are needed to determine its effectiveness.

### **Correlation between Grades and SAGE Scores by Years of Teacher Experience**

Teaching experience is a valuable asset in the educational system (Cleary & Groer, 1994; Sabers, Cushing, & Berliner, 1991; Westerman, 1991). Each year of additional teaching experience provides teachers with more resources to navigate the classroom setting and help students learn. In contrast, newer teachers are often tasked with developing new curriculum and trying to “keep their head above water.” Question four focuses on grades and end-of-level SAGE test correlation in relation to years of teaching experience from both the standards-based grading cohort and the traditional grading group.

The first group we looked at were the teachers with 0-5 years teaching experience. The data indicated that both the standards-based grading cohort and the traditional grading group of teachers assigned grades to students that were higher than what one would expect them to score on the end-of-level SAGE test. However, the standards-based grading cohort grades were statistically and significantly more correlated with SAGE scores than the grades from the traditional group. In fact, the teachers with 0-5 years of teaching experience in the standards-based cohort tied for the highest rate of correlation among all the group sets (Table 11). The teachers with 0-5 years of teaching experience

in the traditional grading group were very consistent with the average correlation among all groups in the traditional grading group.

The next group we examined were teachers with 6-10 years of teaching experience. Results from the analysis showed the standards-based grading cohort to under predict the end-of-level SAGE scores, meaning the grades were actually lower than how the students performed on the SAGE test. The opposite was true for the traditional grading group. The traditional grading group assigned higher grades to reflect student learning than students demonstrated on the end-of-level SAGE test. Even though one group over predicted and the other group under predicted, neither group was very effective in predicting the correlation between grades and SAGE test scores. This means that teachers in the 6-10 years of teaching experience group demonstrated the worst overall correlation between grades and test scores.

The third group observed included teachers with 11-15 years of teaching experience. Both grading groups assigned grades that were more predictive of student SAGE scores than their peers with 6-10 years of teaching experience. However, the same was true for the group with 11-15 years of teaching experience as with the groups who have 6-10 years of teaching experience. The standards-based cohort assigned grades that were lower than the achieved SAGE scores, and the traditional group assigned grades that were higher than the achieved SAGE scores. Data demonstrate the traditional group with 11-15 years of teaching experience to have the 3<sup>rd</sup> highest rate of predictability for grades and SAGE test scores.

Last, teachers with 16 or more years of teaching experience were analyzed.

Teachers in the standards-based grading cohort performed at their worst as related to the predictability of grades and SAGE scores. On the other hand, teachers from the traditional grading group performed at their best. Teachers with 16 years or more of teaching experience in the traditional grading group tied for the highest predictability rate with the standards-based grading teachers who have 0-5 years of teaching experience. This means that teachers become better graders as they gain teaching experience. Additionally, teachers with more experience do not implement change as effectively as their peers with less experience.

Results from question #4 provided a lot of interesting data. The two groups with the highest predictability between grades and SAGE test scores are teachers with the fewest and most years of teaching experience. Teachers who have 0-5 years of teaching experience are new to the teaching profession and do not have a lot of experience with grading. The teachers from the 0-5 years of teaching experience group who participated in the standards-based grading cohort were possibly easier to train because they did not have as many grading habits to break than their peers with more teaching experience. As the standards-based grading cohort had more years of teaching experience, the predictability of grades and SAGE scores worsened, with an exception of the teachers with 11-15 years of teaching experience.

In sum, the teachers in the traditional grading group became better at assigning grades that predict end-of-level tests scores as their years of experience increased. If assigning grades that truly reflect student learning is important, universities and school districts need to focus on training teachers to grade effectively. It is unrealistic in any



field of work to expect an employee to become proficient at a major task, only after the employee has worked for 16 years. If new teachers can be taught to grade effectively during their university studies, and that knowledge can be followed up with productive professional development early in their careers, schools will contain teachers who are effective at assessing student learning and assigning grades. Employing and training teachers to grade effectively will break the cycle of traditional grading and begin a new age for education.

The results of this research study add to the limited research on standards-based grading. While it is difficult to control for all factors, results from this research study indicate that standards-based grading has proven benefits to affect student learning and progress. Educators must continue to examine their current grading methods and determine if how they are assigning grades is an accurate reflection of what a student has learned.

### **Practical Implications of the Study**

Educators are spending a great deal of time, money, and resources to improve the quality of education offered to students. Very little of that time, money, and resources are spent developing and training teachers in grading accurately and effectively. It seems contrary to dedicate so many resources in identifying standards and developing common assessments if those assessments are not going to be graded accurately and systematically across the board.

Education professionals need to dedicate more time and resources to analyze the

research in grading and provide teachers with the necessary professional development and tools to ensure student grades are reflective of what a student knows. Simply continuing to grade student work the same way educators have always graded student work is irresponsible if the grading method does not work (Guskey, 2015). Colleges should either continue or begin to train students earning a degree in education in effective grading methods, and schools should be mentoring new teachers to implement effective grading. If nothing changes, the cycle of grading will continue to persist. If poor practices continue, we will go another 100 years with educators still using the same grading practices designed to motivate students to work in factories and fields.

### **Theoretical Implications of the Study**

Very little research has been completed in the area of standards-based grading. In fact, grading is often an area of education that is overlooked in research areas. There are a variety of data and many variables that factor into an assigned grade, and it can be difficult to determine if the interventions provided in the research are making a difference.

It is important that grading no longer be overlooked in the area of educational research. As education continues to change and improve with newer standards and better technology, the research on effective grading needs to improve as well. Many of the books and articles on grading promote ideas on how to grade effectively, and while those ideas may be good for students, too many of the ideas lack research to support using them. Guskey (2015) recommends schools implement grading practices that are backed

by research. In order for this to be a possibility and reality, increased research on grading needs to be done.

Technology has come a long way to help teachers become more efficient graders. Assignments and assessments posted in an online format can be graded by the computer and results are instantaneous. Quick feedback on assessments can provide data for students and teachers to help determine what students need next.

This research study has shown that assessing and assigning grades can be researched. Standards-based grading can be beneficial in education and should be considered as a grading method for schools to implement. More research and training needs to be completed to determine if standards-based grading is the most effective form of grading and how it could be implemented in all educational settings.

### **Limitations of the Research**

In spite of the researchers' efforts to address all the areas of validity in the research, some limitations remained. We found several limitations as we examined the research output. The first limitation was that the external validity of the study was limited to one school district in the Western U.S. Only sampling student and teacher student data from the same area might not be reflective of schools in other areas of the country. We tried to control this limitation by sampling teachers from different areas of the school district.

Another factor limiting the validity of the study was the selection of teachers to implement standards-based grading. While the standards-based grading cohort of teachers

received the same professional development and training, not all of the teachers implemented standards-based grading with the same consistency as others. In addition, researchers were unable to determine if the positive results in grades and test scores were due to better teaching methods or because of the grading interventions. The researchers attempted to minimize limitation of differing levels of implementation by having teachers in the standards-based grading experiment complete a survey identifying which points of standards-based grading they were implementing effectively. Results from the survey show teachers were implementing standards-based grading with fidelity and consistent with the training they received.

Another limitation of the study includes the survey being a self-report model, and teachers could have minimized or inflated their own scores. It would have been ideal to have observers help determine the rigor with which teachers in the standards-based grading cohort study were implementing the grading practices. Researchers attempted to minimize the limitation of having a self-report model by informing participants that survey results were anonymous and they would not be penalized or rewarded for the answers provided on the survey.

The fourth limitation identified in the study was time. The collected data only spanned one school year. When implementing a new intervention, it is not uncommon to experience a j-curve effect. Results in the beginning of the intervention may dip lower than results would have been without the intervention. Obtaining multiple years of student grades and testing data would have added to the strength of the research and help researchers determine results would continue to improve. A final limitation is the validity

of the end-of-level SAGE tests. While end-of-level testing is a norm in education, the validity of end-of-level tests demonstrating what a student knows is suspect. Utah law allows parents to opt their children out of SAGE tests without penalty. In addition, end-of-level SAGE tests cannot be counted toward a student's grade. With that in mind, some students do not take the SAGE test seriously, producing test results that may not reflect what a student knows. We tried to minimize the limitation by obtaining data from thousands of students, but the limitation still exists.

### **Recommendations for Future Research**

After reviewing the results and making conclusions of the research study, we have determined three areas that need additional research. Within each of the three areas of need, we provide recommendations for addressing the need. The three areas include replicating the research study on a larger scale, completing a case study of teachers implementing standards-based grading, and researching other grading methods to determine their effectiveness.

While the research study provided excellent results and items for discussion, the limitations of using data from the same school district over a one school-year time span need to be enlarged. With that, we feel the study should be replicated on a larger scale. Many schools and school districts across the nation have implemented standards-based grading. The data from these schools need to be analyzed to determine the benefits and usage of standards-based grading.

Implementing a new grading method can be taxing on a teacher. Parents are used

to the grading methods teachers used when they were in school. Students are used to the grading methods in other classes. Researchers looking at a case study of a small group of teachers implementing standards-based grading would provide needed research and data for other educators to observe and learn from. If a study is able to provide research detailing what worked and did not work in a classroom implementing a grading change, it would help many educators avoid the same pitfalls and prevent them from going back to traditional grading practices when they encounter issues.

The research on grading, not just standards-based grading, needs to increase. If educational experts are going to identify an effective way to assess and grade students, all grading methods need to be researched. The research on standards-based grading is minimal, but educational speakers have written several books and performed a great deal of professional development across the country. Schools need to experiment with grading methods and determine the effectiveness of each grading method supported with research because as research grows, educators and students are able to benefit.

### **Conclusion**

As education reforms and adapts to meet the needs of our students, reporting student learning and progress through the use of grades must do the same. It is recommended educators work toward finding the most appropriate way to represent student learning. Improved technology allows parents and students to see up-to-the-minute grades. If those grades do not help stakeholders identify areas of strengths and weaknesses, then it only adds to the confusion of what a student has or has not learned.

In the study, we found several key items that will add to the research on grading that already exists:

Students who attended a standards-based grading class earned higher grades and performed better on the end-of-level SAGE test. Earning higher grades provides increasing opportunities for students to play on athletic teams, be on honor rolls, earn college admission, and increase their chances for college scholarships. Students who show higher proficiency on high-stakes tests increase their chances of being accepted into college and earning a scholarship. Colleges use grades and high-stakes test scores to determine college admission and scholarship opportunities.

Students who attended a standards-based grading class showed higher learning growth as measured by the end-of-level SAGE test. Students are expected to grow a grade level after each year of instruction. As noted in the USDE (2009) Race to the Top program, effective teachers can help students grow more than one grade level in a school year. If changing the grading practices creates opportunities for increased learning, educators need to jump on board and do what is most beneficial for students.

The correlation between assigned grades and end-of-level SAGE test scores was not better or worse by being in a standards-based or traditional graded classroom. Ideally, grades and test scores would be correlated with each other. As educators receive more professional development and training, grading practices will improve and accurately reflect what a student knows and can do. Being able to accurately determine a student's learning and progress at any given point in the school year is beneficial for all educational stakeholders. Teachers would be able to determine which students need extra

support and which students are ready to move on. Parents could provide targeted support to students as they try to meet standards. Lastly, students know exactly which standards they are proficient in and which standards they need to focus on.

End-of-level SAGE test scores were higher than the grades predicted in the standards-based grading cohort. Students in the standards-based cohort were tasked with ensuring proficiency on standards throughout the school year. Since the end-of-level SAGE test was based on those standards learned during the school year, students were prepared to perform well on the standards-based test.

End-of-level SAGE test scores were lower than the grades predicted in the traditional grading group. In traditional grading, student grades are commonly inflated with extra credit or points assigned for student behavior, participation, and attendance. Students in the traditional grading group were assigned grades that were higher than what they demonstrated on the end-of-level SAGE test. Assigning credit for participation, behavior, and attendance is not a true reflection of what a student knows.

Teachers with the least amount of experience (0-5 years) from the standards-based grading cohort were equivalent in accuracy with teachers who had the most amount of experience (16+ years) from the traditional grading group at assigning grades predictive of end-of-level SAGE test scores. Two things could have factored into the success of teachers with 0-5 years' experience: universities are working with teacher education students on grading practices or new teachers had not developed grading habits and they were easier to train. Teachers with more experience had a harder time breaking traditional grading habits. Universities and educational leaders must begin to provide



training on effective grading practices. In my experience, new teachers would like to learn effective grading practices but often do not know where to turn for help. If we do not make a change in teacher training, we will not see a change in grading practices. Experience in education is a valuable asset to have; however, we cannot wait 16 years or more for teachers to become proficient in assigning grades. Educational leaders need to do a better job of ensuring teachers are proficient at assigning grades and using effective grading techniques.

While this study contributed to the research on grading and standards-based grading, more research is needed to determine what grading method is most effective. More effort and research from grading researchers will help educators understand the complexity and benefits of a grading reform. Grading experts need to support their grading recommendations with research from the field. Our research study is only a small piece to the puzzle toward effective grading. We need more research to truly understand where education needs to go next.

The purposes of grading must be at the forefront of any effective grading system (Airasian, 2001; Feldmesser, 1971; Frisbie & Waltman, 1992; Guskey & Bailey, 2001; Linn, 1983). The six major categories that define grading practices are as follows.

1. Communicate information about student achievement to parents and others.
2. Provide information and feedback to students for self-evaluation.
3. Select, identify, or group students for specific educational paths and programs. This could include higher level AP and honors classes or possibly remedial and special education classes.
4. Provide incentive and motivation for students to learn.
5. Evaluate the effectiveness of instructional programs. Are students learning

and is the content valuable?

6. Provide evidence of students' lack of effort or inappropriate responsibility.

Each of these purposes are important to reflect on as grading methods and practices are considered. If standards-based grading is implemented as designed, all six of the grading purposes can be met effectively. Educators and researchers must look at current practices and make the necessary adjustments to meet the needs of students

The research on grading continues to grow and progress. Teachers and educational leaders need to take the steps necessary to improve the accuracy of assigning grades. As grading improves, the conversations between stakeholders and educators will improve. Instead of asking, "How can my child earn an A?" Stakeholders will be asking, "What can my child do to learn this standard?" When the focus is on learning instead of grading, the true goal of education will be met.

## REFERENCES

- Airasian, P. W. (2001). *Classroom assessment: Concepts and applications* (4<sup>th</sup> ed.). New York, NY: McGraw-Hill.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bancroft, K. (2010). Implementing the mandate: The limitations of benchmark tests. *Educational Assessment, Evaluation and Accountability*, 22(1), 53-72.
- Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research and Development*, 18(1), 57-75.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.
- Bobbit, F. (1912). The elimination of waste in education. *The Elementary School Teacher*, 12, 259-271.
- Bobbit, F. (1924). *How to make a curriculum*. Boston, MA: Houghton Mifflin.
- Bond, L. A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2), 1-3.
- Brennan, R., Kim, J., Wenz-Gross, M., & Siperstein, G. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review*, 71(2), 173-217.
- Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research and Evaluation*, 16(17), 1-12.
- Brookhart, S. M. (1993). Teachers' grading: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, NJ: Pearson Education.
- Brookhart, S. M. (2011). Starting the conversations about grading. *Educational Leadership*, 69(3), 10-14.
- Brookhart, S. M. (2012). Grading. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257-71). Thousand Oaks, CA: Sage.

- Buckles, S., Schug, M. C., & Watts, M. (2001). A national survey of state assessment practices in the social studies. *The Social Studies*, 92(4), 141-146.
- Canady, R. L. (n.d.). *Grading practices that increase/decrease the odds for student success: General grading information packet*. Retrieved from <http://robertlynnncanady.com/canady /grading/Grading-Packet-10-11.pdf>
- Canady, R. L., Canady, C. E., & Meek, A. (2017). *Beyond the grade: Refining practices that boost student achievement*. Bloomington, IN: Solution Tree Press.
- Cleary, M. J., & Groer, S. (1994). Inflight decisions of expert and novice health teachers. *Journal of School Health*, 64, 110-114.
- Cornelius, K. E. (2013). Formative assessment made easy: Templates for collecting daily data in inclusive classrooms. *Teaching Exceptional Children*, 45(5), 14-21.
- Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, 40(1), 67-87.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Macmillan.
- Dueck, M. (2014). *Grading smarter, not harder: Assessment strategies that motivate kids and help them learn*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Dufour, R., Dufour, R., Eaker, R. & Many, T. (2010). *Learning by doing: A handbook for professional learning communities at work*. Bloomington, IN: Solution-Tree Press.
- Duncan, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research* 53(1), 1-21.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.
- Elmore, R., Abelman, C., & Fuhman, S. (1996). The new accountability in state education reform: From process to performance. In H. Ladd (Ed.), *Holding schools accountable* (pp. 65-98). Washington, DC: Brookings Institution Press.

- Feldman, A., Kropf, A., & Alibrandi, M. (1998). Grading with points: The determination of report card grades by high school science teachers. *School Science and Mathematics*, 98(3), 140-148.
- Feldmesser, R. A. (1971, February). *The positive functions of grades*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. New York, NY: McGraw-Hill International.
- Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice*, 12(3), 23-30.
- Friedman, S. J., & Troug, A. J. (1998). Evaluation of high school teachers' written grading policies. *ERS Spectrum*, 17(3), 34-42.
- Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practices*, 11(3), 35-42.
- Griffin, T. J., Hilton, J., III, Plummer, K., & Barret, D. (2014). Correlation between grade point averages and student evaluation of teaching scores: Taking a closer look. *Assessment & Evaluation in Higher Education*, 39(3), 339-348.
- Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership*, 59, 20-27.
- Guskey, T. R. (2002). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan*, 83(10), 775-780.
- Guskey, T. R. (2004). 0 alternatives. *Principal Leadership*, 5(2), 49-53.
- Guskey, T. R. (2008). *Practical solutions for serious problems in standards-based grading*. Thousand Oaks, CA, Corwin.
- Guskey, T. R. (2015). *On your mark: Challenging the conventions of grading and reporting*. Bloomington, IN: Solution Tree Press.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin.
- Guskey, T. R., & Bailey, J. M. (2010). *Developing standards-based report cards*. Thousand Oaks, CA: Corwin.

- Haladyna, T. M. (1999). *A complete guide to student grading*. Boston, MA: Allyn & Bacon.
- Hamilton, L., Stecher, B., & Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions*. Santa Monica, CA: Rand Education.
- Hanushek, E. A., & Raymond, M. E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2, 406-415.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hirst, B. A., Jr. (1973). *A project for the development of a multi-state consortium for the production of performance objectives and criterion measures in occupational education. Second report*. Retrieved from <http://www.eric.ed.gov/fulltext/ED133515.pdf>
- Hojnoski, R. L., Gischlar, K. L., & Missall, K. N. (2009). Improving child outcomes with data-based decision making: Collecting data. *Young Exceptional Children*, 12(3), 32-44.
- Hooper, J., & Cowell, R. (2014). Standards-based grading: History adjusted true score. *Educational Assessment*, 19(1), 58-76.
- Howley, A., Kusimo, P. S., & Parrot, L. (2000). Grading and the ethos of effort. *Learning Environments Research* 3(3), 229-246.
- Iamarino, D. L. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education*, 17(2), 1-10.
- Jensen, E. (2009). *Teaching with poverty in mind: What being poor does to kids' brains and what schools can do about it*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. New York, NY: Psychology Press.
- Lalley, J. P., & Gentile, J. R. (2009). Classroom assessment and grading to assure mastery. *Theory into Practice* 48(1), 28-35.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.

- Linn, R. L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, 20(2), 179-189.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2006). *Classroom assessment & grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2011). *Formative assessment & standards-based grading*. Bloomington, IN: Solution Tree Press.
- Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership*, 69(3), 34-39.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Martinez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, 14, 78-102.
- McMillan, J. H. (2012). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 3-16.). Thousand Oaks, CA: Sage.
- McMunn, N., Schenck, P., & McColskey, W. (2003, April). *Standards-based assessment, grading, and reporting in classrooms: Can district training and support change teacher practice?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED475763)
- McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63, 10-17.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*. Washington DC: Authors.

- Newton, J., & Winches, B. (2013). How did they maximize learning for all of those students? *Reading Improvement*, 69(2), 71-74.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- O'Connor, K. (2002). *How to grade for learning: Linking grades to standards* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Corwin
- O'Connor, K. (2011). *A repair kit for grading: 15 fixes for broken grades* (2<sup>nd</sup> ed.). Boston, MA: Pearson.
- O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership*, 69(3), 40-44.
- Phelps, R. P. (2012). The effect of testing on student achievement, 1910-2010. *International Journal of Testing*, 12(1), 21-43.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26, 1372-1380.
- Ravitch, D. (1995). *National standards in American education*. Washington DC: Brookings Institution Press.
- Raymond, G. (2013). The trouble with the curve: An argument for the abolishment of norm-referenced evaluation. *i-Manager's Journal on English Language Teaching* 3(1), 1-9.
- Reeves, D. B. (2004). *Making standards work: How to implement standards-based assessments in the classroom, school, and district*. Englewood, CO: Advanced Learning Press.
- Reeves, D. B. (2010). *Standards, assessment, and accountability: Real questions from educators with real answers from Douglas B. Reeves, Ph.D.* Englewood, CA: The Leadership and Learning Center.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education*, 12, 315-325.
- Rshaid, G. (2011). *Learning for the future: Rethinking schools for the 21<sup>st</sup> century*. Englewood, CO: Lead+Learn Press.
- Sabers, D. S., Cushing, K. S., & Berliner, D.C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, 28, 63-88.



- Scarino, A. (2005). Teacher judgments: Going beyond criteria for judging performance. *Babel*, 39(3), 8-16.
- Schimmer, T. (2016). *Grading from the inside out: Bringing accuracy to student assessment through a standards-based mindset*. Bloomington, IN: Solution Tree Press.
- Schneider, M. C., Egan, K. L., & Julian, M. W. (2012). Classroom assessment in the context of high-stakes testing. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 55-70). Thousand Oaks, CA: Sage.
- Scriffiny, P. L. (2008). Seven reasons for standards-based grading. *Educational Leadership*, 66(2), 70-74.
- Shapiro, E. S., Hilt-Panahon, A., Gischlar, K. L., Semeniak, K., Leichman, E., & Bowles, S. (2012). An analysis of consistency between team decisions and reading assessment data within an RTI model. *Remedial and Special Education*, 33(6), 335-347. doi: 0741932510397763
- Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4<sup>th</sup> ed., pp. 1066-1101). Washington, DC: American Educational Research Association.
- Small, A. W. (1896). Demands of sociology upon pedagogy. *Journal of Proceedings and Addresses of the Thirty-Fifth Annual Meeting of the National Education Association*, 174-184.
- Smith, C. W. (1973). *Criterion-referenced assessment*. Retrieved from files.eric.ed.gov/fulltext/ED081843.pdf
- Smith, M. I., Heinecke, W., & Noble, A. J. (1999). Assessment policy and political spectacle. *The Teachers College Record*, 101(2), 157-191.
- Starch, D., & Elliot, E. C. (1912). Reliability of the grading of high school work in English. *School Review*, 20(7), 442-457.
- Starch, D., & Elliot, E. C. (1913). Reliability of the grading of high school work in mathematics. *School Review*, 21(4), 254-259.
- Stiggins, R. J., & Chappuis, J. (2011). *An introduction to student-involved assessment for learning* (6<sup>th</sup> ed.). Boston, MA: Pearson Education.

- Sturgis, C., Patrick, S., & Pittenger, L. (2011). *It's not a matter of time: Highlights from the 2011 Competency-Based Learning Summit*. Vienna, VA: International Association for K-12 Online Learning.
- Suurtamm, C. (2004). Developing authentic assessments: Case studies of secondary school mathematics teachers' experiences. *Canadian School of Science, Mathematics and Technology Education*, 4(4), 497-513.
- Thomas, J. Y., & Brady, K. P. (2005). The elementary and secondary education act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education*, 29, 51-67.
- Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: That's my grade. That's me., *Educational Psychologist*, 32(2), 107-123.
- Tomlinson, C. A., & McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. Alexandria, VA: Association for Supervision and Curriculum Development.
- U.S. Department of Education. (2009). *Race to the top executive summary*. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-/summary.pdf>
- Utah State Senate. (2013). *Utah school grading*. Retrieved from <http://utahschoolgrading.com/faq/>
- Vatterott, C. (2009). *Rethinking homework: Best practices that support diverse needs*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Walker, C. A. (2002). Saving your rookie managers from themselves. *Harvard Business Review*, 80(4), 97-111.
- Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223-240.
- Welsh, M. E., D'Agostino, J. V., & Kaniskan, B. (2013). Grading as a reform effort: Do standards-based grades converge with test scores? *Educational Measurement: Issues and Practice*, 32(2), 26-36.
- Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education*, 42, 292- 305.
- Wiggins, G. P. (1993). *Assessing student performance*. San Francisco, CA: Jossey-Bass.
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39, 1-37.

## APPENDICES

## Appendix A

### Standards-Based Grading in the Classroom

**Standards-Based Grading in the Classroom**

1. Last Name:
2. First Name:
3. What school do you teach at?
4. What subject do you teach?
  - a. Science
  - b. English
  - c. Math
5. I have identified learning targets that demonstrate what students should know and be able to do.
  - a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree
  - e. Disagree
  - f. Strongly disagree
6. I connect each assessment to specific learning targets.
  - a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree
  - e. Disagree
  - f. Strongly disagree
7. I have validated my learning targets. (choose all that apply)
  - a. in my classroom
  - b. with teachers at my school who teach common courses
  - c. with teachers throughout the district
  - d. I have not validated my learning targets

8. I encourage students to retake assessments.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

9. When students retake assessments what type of penalty is given?

- a. No penalty
- b. 5%-25% penalty
- c. 26-50% penalty
- d. More than 50% penalty
- e. I do not allow students to retake assessments

10. I allow students to complete extra credit to raise their grade.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

11. I average all of the scores a student has earned to calculate a grade.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

12. When determining grades, I put more emphasis on recent evidence rather than older evidence.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

13. I regularly communicate and provide learning targets for students.
- a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree
  - e. Disagree
  - f. Strongly disagree
14. Attendance, student participation, and behavior are included in the letter grades I assign to students.
- a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree
  - e. Disagree
  - f. Strongly disagree
15. What percentage does homework count for when determining a student's grade in your classroom?
- a. 0%
  - b. 10% or less
  - c. 11%-25%
  - d. 26%-50%
  - e. More than 50%
16. Students, parents, and other educational stakeholders can look at grades assigned by me and identify which standards a student is proficient and deficient in.
- a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree
  - e. Disagree
  - f. Strongly disagree
17. Full credit should not be given to students who hand in late work.
- a. Strongly Agree
  - b. Agree
  - c. Somewhat agree
  - d. Somewhat disagree

- e. Disagree
- f. Strongly disagree

18. In my class, the grading system communicates what students know and are able to do, rather than reward student accomplishment.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

19. In my class, if a student has a missing assessment, I assign a zero to the student's grade.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree

20. My assessments are criterion-referenced or evidence-based, measuring a certain level of achievement rather than comparative to other student work.

- a. Strongly Agree
- b. Agree
- c. Somewhat agree
- d. Somewhat disagree
- e. Disagree
- f. Strongly disagree



## Appendix B

### GPA and SAGE Error Bar Charts

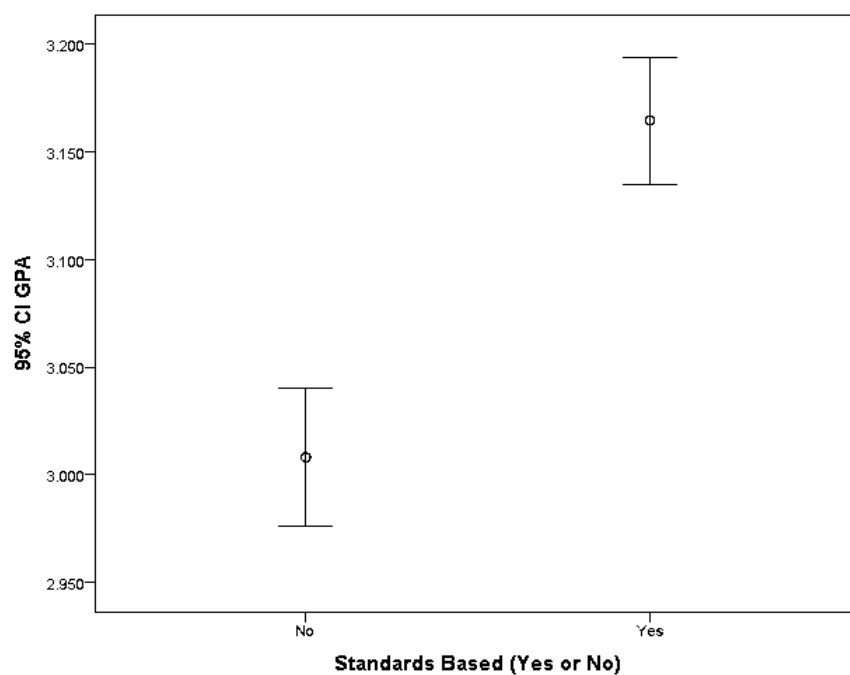


Figure B1. GPA analysis.

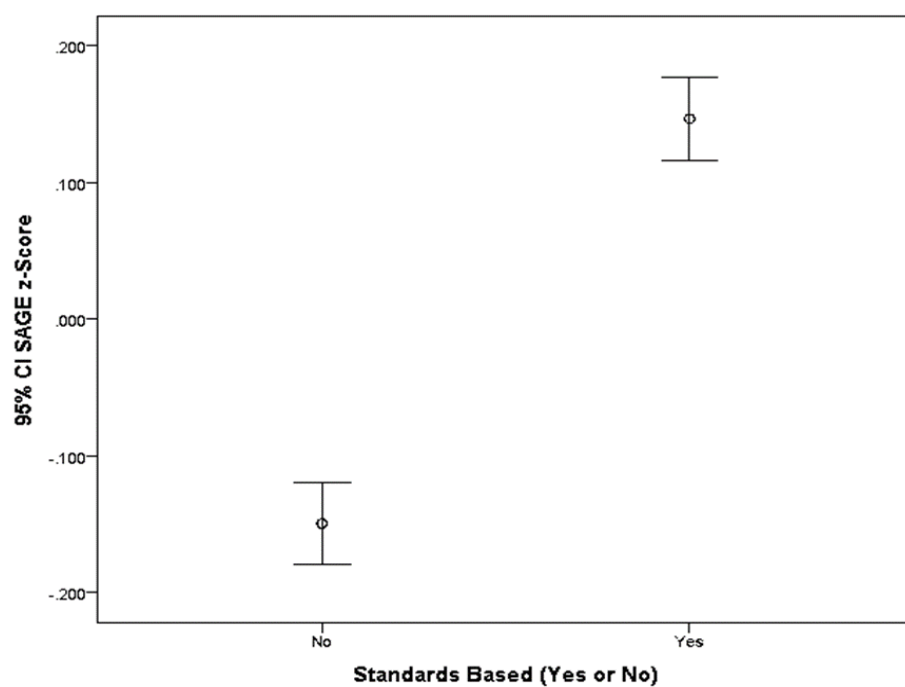


Figure B2. SAGE  $z$  score analysis.

## Appendix C

### SAGE $z$ Score Regression Residuals

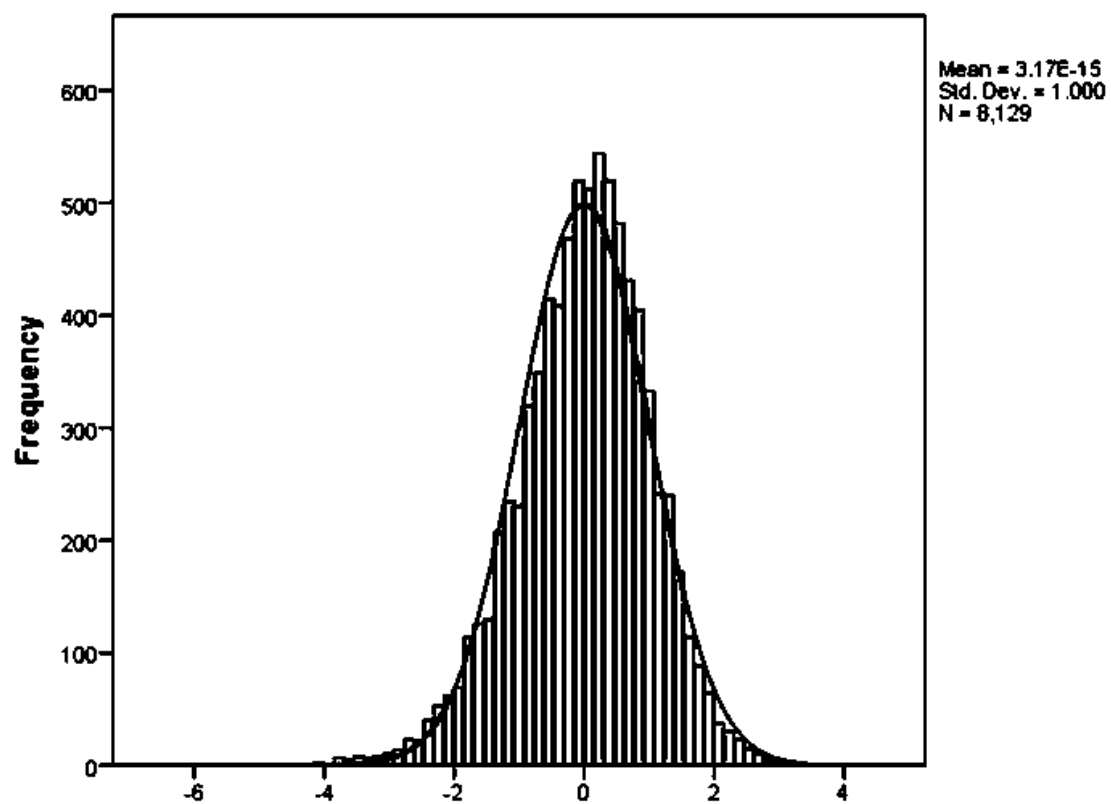


Figure C1. SAGE  $z$  score regression residual.

## Appendix D

### Error Bar Chart Showing Confidency Level for Residuals

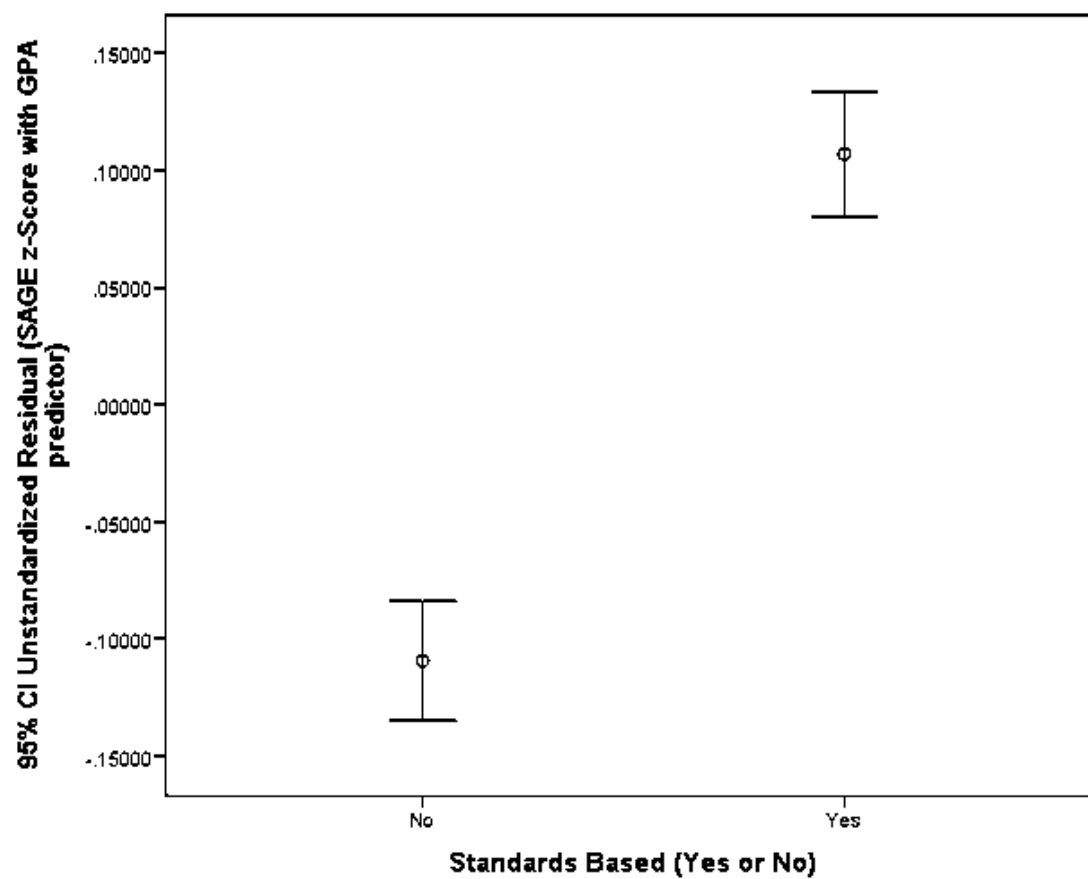


Figure D1. Error bar chart: Confidence level for residuals.

## Appendix E

### Bar Graph Showing SGP Score Analysis

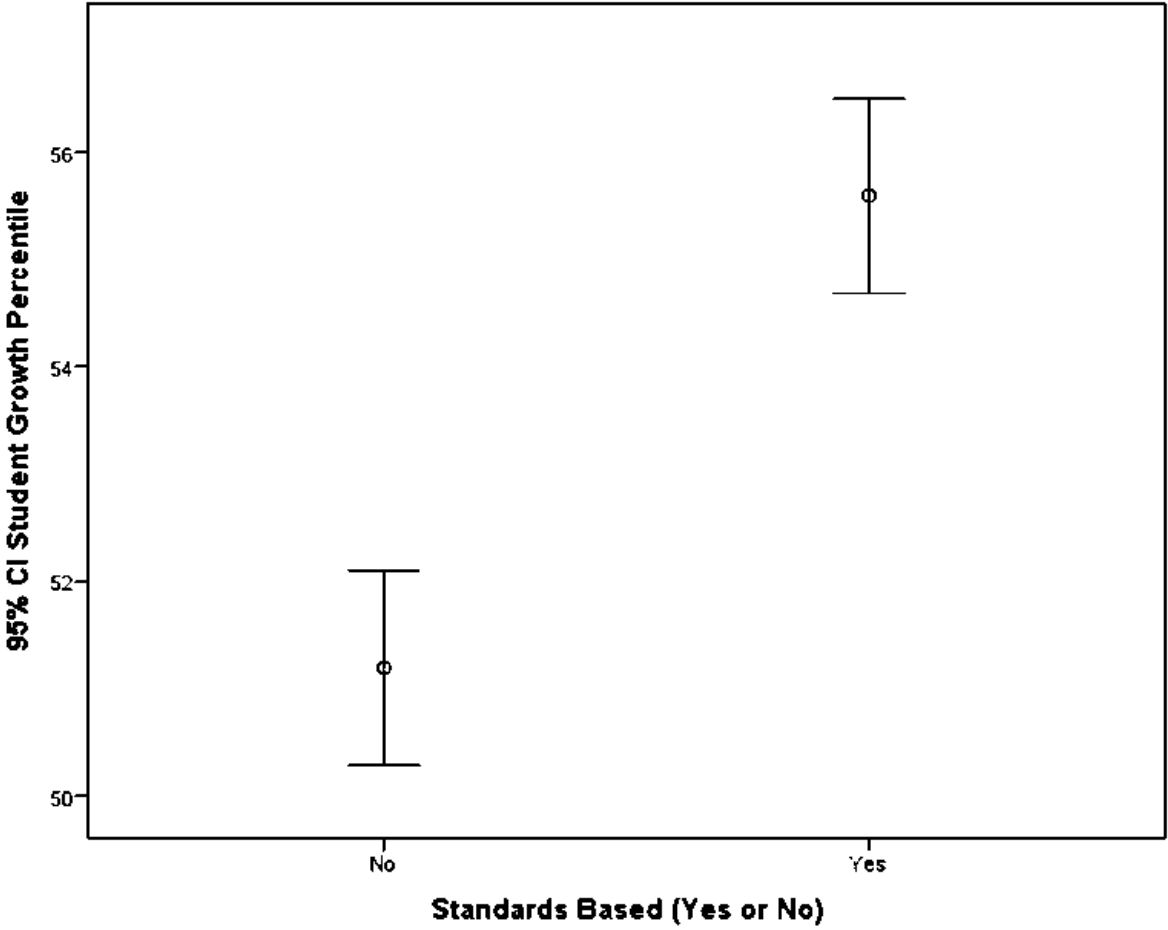


Figure E1. SGP score analysis.



## Appendix F

### Error Bar Charts by Content

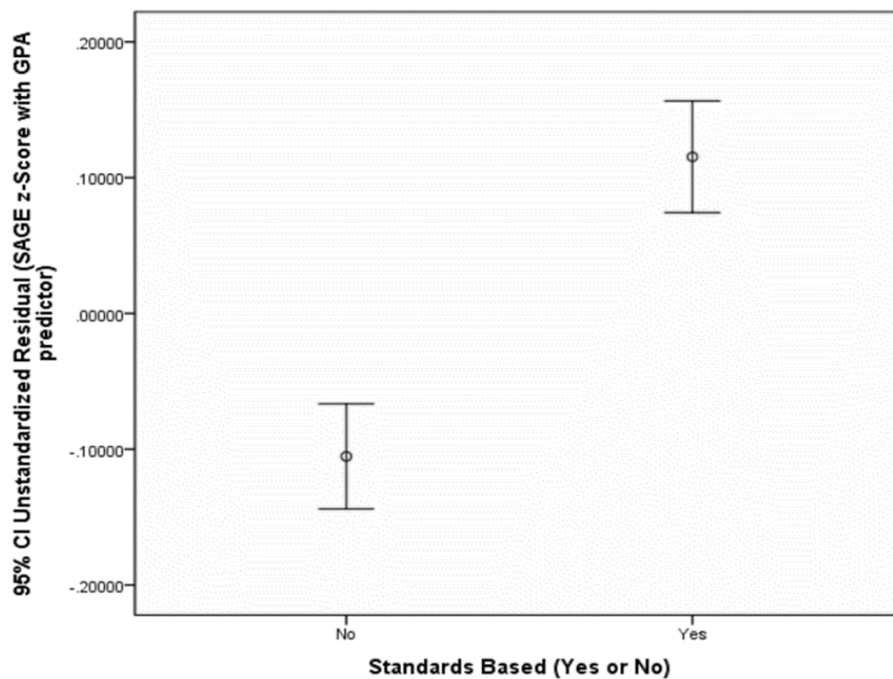


Figure F1. Language arts error bar chart: Confidence level for residuals.

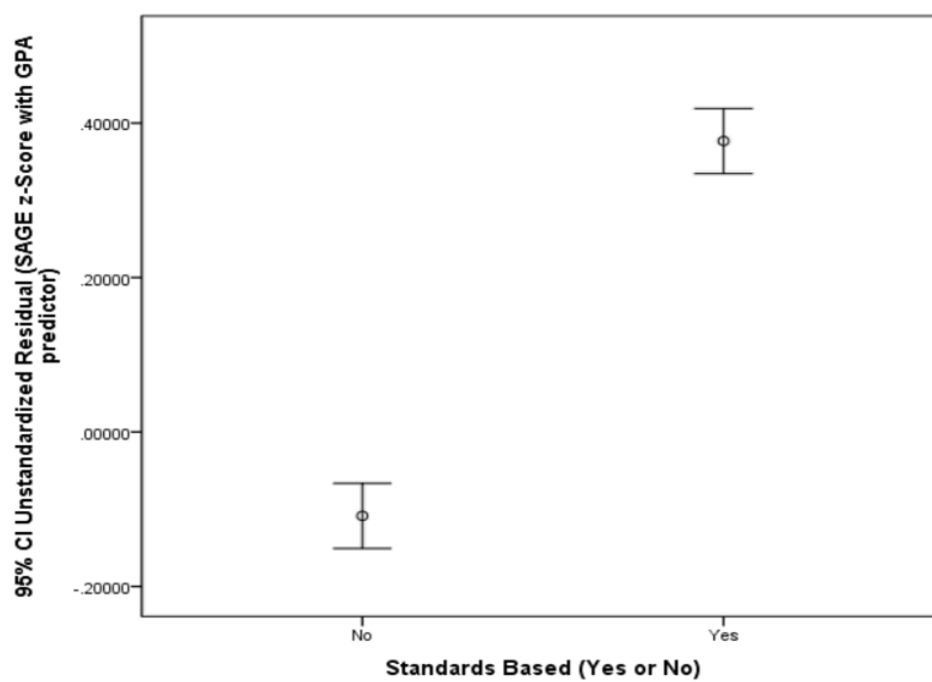


Figure F2. Math error bar chart: Confidence level for residuals.

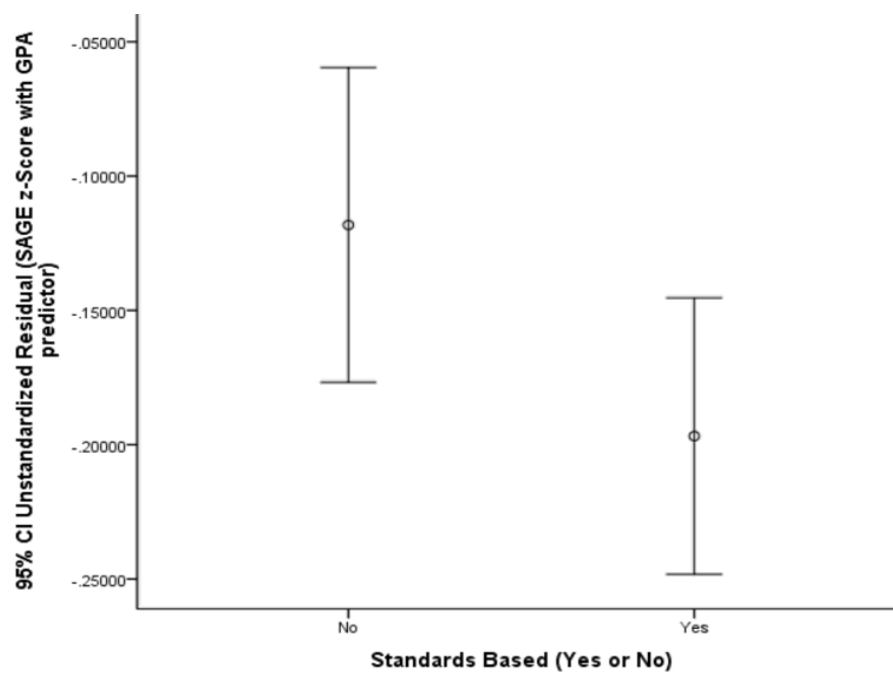


Figure F3. Science error bar chart: Confidence level for residuals.

## Appendix G

### Error Bar Charts by Experience

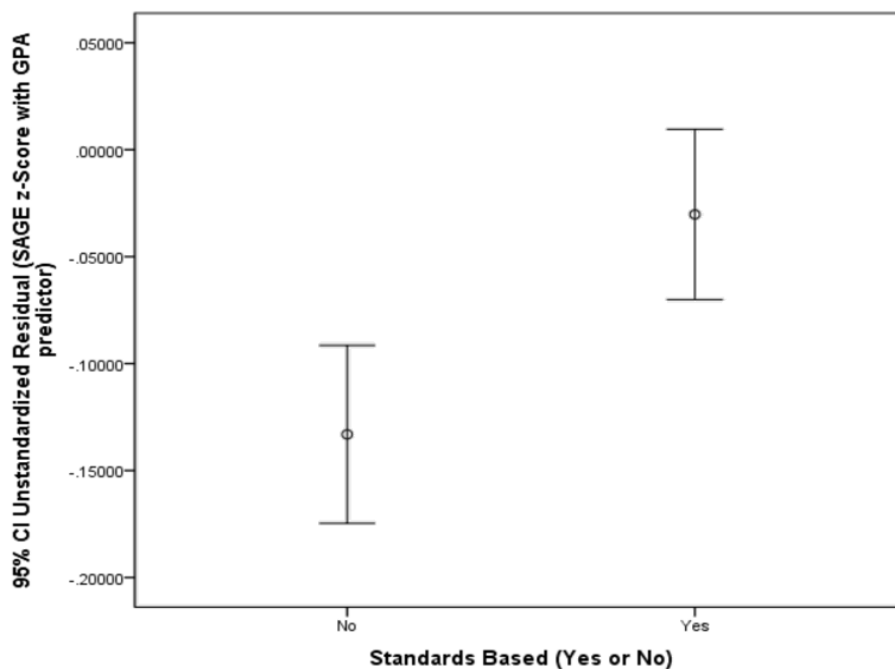


Figure G1. 0-5 years teaching experience error bar chart: Confidence level for residuals.

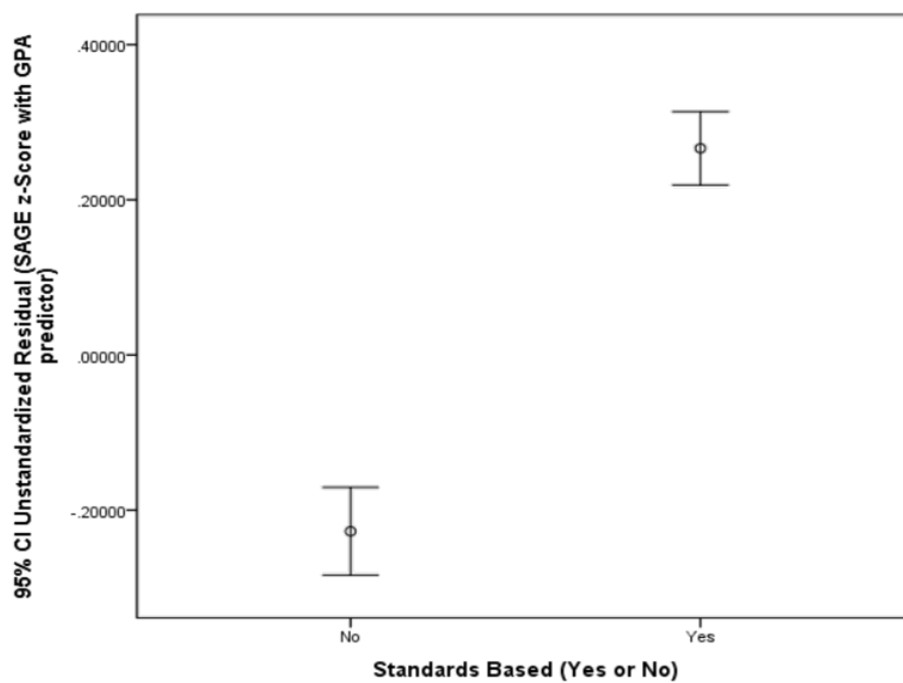


Figure G2. 6-10 years teaching experience error bar chart: Confidence level for residuals.

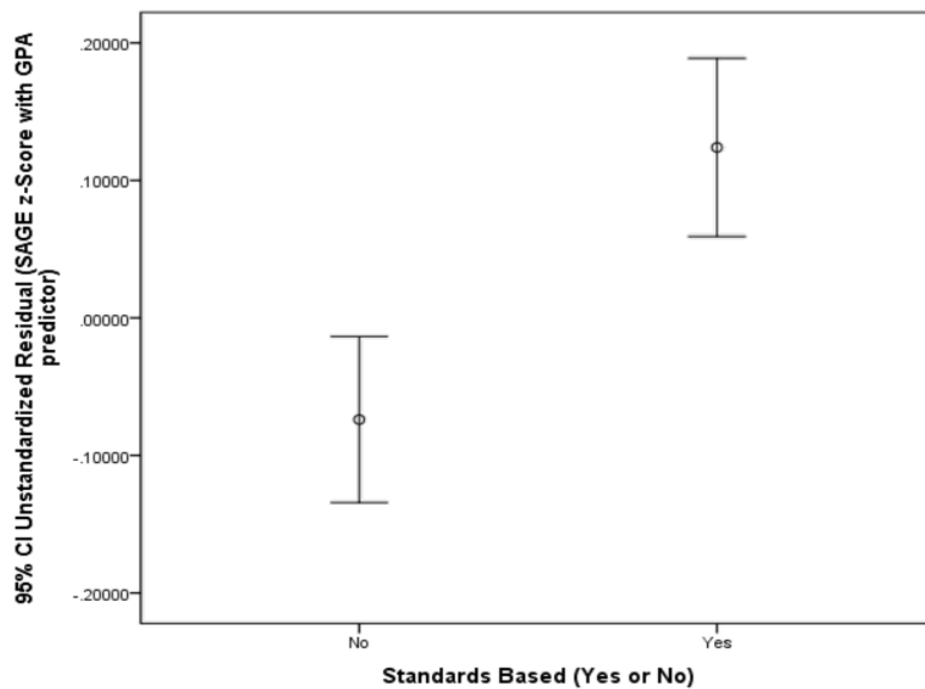


Figure G3. 11-15 years teaching experience error bar chart: Confidence level for residuals.

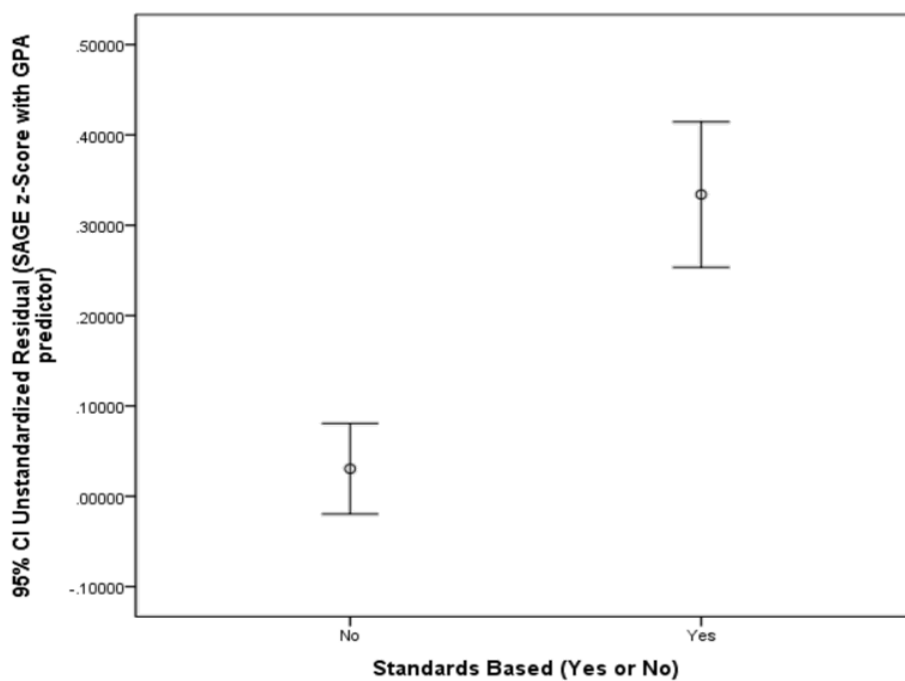


Figure G4. 16+ years teaching experience error bar chart: Confidence level for residuals.

## CURRICULUM VITAE

TYLER R. POLL

## ADMIN

7902 S. 2450 E., South, Weber, UT 84405 | 801-458-7740 | [tpoll@dsdmail.net](mailto:tpoll@dsdmail.net)

## EDUCATION

**Doctor of Philosophy****2019**

Utah State University

Area of Concentration: Instructional Leadership

Dissertation: "Standards-Based Grading: A Correlational Study Between Grades and End-of-Level Test Scores"

**Administrative and Supervisory Endorsement****2013**

Utah State University

Area of Concentration: Instructional Leadership

**Masters of Education****2011**

Weber State University

Area of Concentration: Curriculum and Instruction

**Bachelors of Science****2010**

Weber State University

Area of Concentration: Business Education

Minor: Marketing Education

**Bachelors of Science****2007**

Weber State University

Area of Concentration: Technical Sales

## EXPERIENCE IN EDUCATION

**Principal****2018-Current**

North Davis Junior High

**Assistant Principal****2016-2018**

Northridge High School

**Assistant Principal** **2015-2016**  
North Davis Junior High

**Assistant Principal** **2013-2015**  
Syracuse Junior High

**Career and Technical Education Business and Marketing Teacher** **2010-2013**  
North Davis Junior High

#### RELATED EXPERIENCE

**AdvancEd School Accreditation Team Member** **2014-2018**

**Head Boys Basketball Coach** **2011-2013**

#### PUBLICATIONS AND PAPERS

*“Are They Using the Data? Teacher Perceptions of, Practices with, and Preparation to Use Assessment Data”*  
International Journal of Education **2016**

#### PRESENTATIONS

*“What’s the Problem: Standards-Based Grading in Secondary Schools* **2017**  
Utah Association of Secondary School Principals Conference

*“Grading in 21<sup>st</sup> Century Education”* **2017**  
Davis Education District and School Leadership

#### MEMBERSHIPS

Utah Association of Secondary School Principals  
Davis Education Association  
National Education Association