Summer 2017

What Can One Learn By Measuring Assessments Two Ways: Traditional Grading Scale Versus Standards-Based Grading Method?

Lynn Suzanne Gay
Hamline University

WHAT CAN ONE LEARN BY MEASURING ASSESSMENTS TWO WAYS: TRADITIONAL GRADING SCALE VERSUS STANDARDS-BASED GRADING METHOD?

By

Lynn Suzanne Gay

A capstone submitted in partial fulfillment of the

requirements for the degree of Master of Arts in Teaching

Hamline University

St. Paul, Minnesota

September, 2017

Primary Advisor: Jeffery Fink

Secondary Advisor: Heather Richins

Peer Reviewer: Jodie Bias

To my amazing family, friends, and advisors for all of your support and encouragement.

TABLE OF CONTENTS

CHAPTER ONE INTRODUCTION OF CAPSTONE

Introduction

Assessments are a necessity in the field of education, and the grading of these assessments often feels very personal to both teachers and students. Students regularly express the desire to gain a few more partial credit points, and we as educators balance our attempt to fairly grade students across multiple sections and classes with our desire to see our students become as successful as possible. Standards-based grading has become a popular topic recently, and many school districts are making the change to this new system from traditional grading systems. Having used both methodologies to assess students, I see that each has their merits and deficits. That led me to ask the question: if assessments are graded using a standards-based scale, does that have an impact on the overall grade a student achieves? More specifically, what can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?

In this chapter, I will describe my experiences with both traditional and standards-based grading systems for assessments. I explain from where my interest in the topic originates and the questions that I am hoping to address through my research.

Background of the Researcher

As a student, I was constantly concerned about my grades because I knew that a high grade point average was necessary to attend a "good" university. I decided early in

high school that I wanted to pursue a career in Mechanical Engineering, so of high importance to me were my grades in my mathematics classes. The high school I attended used a traditional grading system, and teachers awarded partial credit if students used a correct process but made mistakes and got an incorrect answer. I recall regularly attempting to convince teachers to award me one or two more points so I could get the highest grade possible. In college, my mathematics and science classes were the same, and we were awarded partial credit for having some correct work even if we did not find the correct answer. Again, my peers and I often spoke with our professors to try to convince them to award additional points so that we could achieve higher grades.

Fast-forward through ten years of engineering jobs and the completion of my classes at Hamline University to receive my teaching license, and now the responsibility fell on me to grade students' work. In my first teaching position, I embarked with my colleagues in the launching year of a new charter school. As the only mathematics teacher, it was incumbent upon me to determine both the curriculum and the grading requirements. Having only known traditional grading systems including the awarding of partial credit on incorrect answers, that system is what I set up at this new school. It worked well, and I was yet unaware that other methods existed. After a rather tumultuous first year teaching in a school that barely survived to see a second year, I made the decision to secure a new position in a public school district that was much more established and stable.

I was relieved and grateful to finally work with colleagues who taught the same classes as me whom I could ask for advice and with whom I could collaborate. I also was introduced to a standards-based grading system at this new school, Mapledale High

School (pseudonym). As my colleagues explained the grading process and system, several aspects made me feel apprehensive. Each test was 9-17 questions that were written at different difficulty levels. The Level 1 questions were basic questions that, if completed correctly, demonstrated a baseline understanding of the concepts. The Level 2 questions asked students to apply the knowledge they learned in the unit and understand the topics more in-depth. The Level 3 question (there was only one) required students to apply concepts from many different units into one question, considered by the teachers to be three to four levels above the current course. The school also had a minimum grade of 50% if a student gave an honest attempt at every question. As we graded these questions, we would mark them either correct or incorrect with no partial credit awarded.

I took issue with all three of these areas as they were explained to me. I was uncomfortable testing students on problems that were unfamiliar to them and required them to apply the knowledge from the unit. In my experiences as a student, if my peers and I had seen a problem unlike any that we had encountered on the homework, we would have accused the teacher of being unfair. In fact, students avoided some of the professors at the university I attended for my undergraduate degree, because those professors asked such difficult and unfamiliar questions on tests, and students generally scored lower grades in those classes. I did not feel comfortable asking my students to do something that frustrated me as a student. As I grew more comfortable at Mapledale High School and with this standards-based grading system, I realized that these very difficult problems allowed students to score bonus points on their tests. I saw many more students attempt problems that they might not normally have been willing to try because they had the chance to achieve higher than 100%.

The second area where I had some question was surrounding the minimum grade of 50%. I have had students that have earned lower grades than that even with partial credit, so were we really doing what was best for students by giving them a minimum grade of 50%? It really made me question if the grades overall would be inflated because the low scores might not be as low as in a traditional grading system environment. However, looking at a traditional grading scale, it is far more heavily weighted toward the failing grades. Only 10% of a possible 100% is set aside for each grade A, B, C, and D for a total of 40%. That means the remaining 60% of a traditional grading scale is considered failing. As I thought about this I wondered if, instead of grades being inflated with standards-based grading as I originally thought, were grades in the traditional system causing a lower average grade in most of our classes?

Finally, I found it difficult to get used to marking problems either correct or incorrect with no partial credit. Having often implored teachers to raise my grade one or two points because I, "knew what I was doing, I just made a silly mistake," I felt as though I was harshly punishing students for making small errors in their work. This is, perhaps, the area in which I learned the most as I was asked to grade in this standards-based system. First, grading became much easier and faster. I finished grading tests within an hour after giving them, so the students could often see their grades online that night. This immediate feedback allowed them to more effectively make sense of their work and apply it to future study. More importantly, however, I had an epiphany. A student's grade was no longer based on my perception of the amount of partial credit he or she deserved. It was based purely on whether or not the student was able to correctly work out and answer a question. It removed all of the subjectivity from grading. When I

chose to become a math teacher, I did so because math almost always has a right answer. I appreciate the black and white nature of a mathematical problem, and that an individual's opinion does not affect their answer. However, when grading on the traditional system and awarding partial credit, I found that I was using my opinion every time I gave a student partial credit on a problem. In other words, it was my opinion that factored into a student's grade instead of purely his or her knowledge of the concepts.

The next year Mapledale High School (pseudonym) did not grow quickly enough to open a full time position for me, so I moved to a new high school again, Stanhope High School (pseudonym). Stanhope employed the traditional grading system, complete with partial credit awarded on problems. As I once again familiarized myself with they system that I had known during my years as a student and first year as a teacher, I wondered if the traditional grading system is really the one that is best for students. I asked myself which system better assessed student knowledge and if one system was in fact superior, which led me to ponder in which grading system are students' grades reflected more accurately? This question was far too broad to answer with any sort of research, but it led to the idea that I would like to look at the differences between traditional grading and standards-based grading of assessments and determine their merits. Specifically, how do students' grades compare when grading with a traditional system versus a standards-based system? While test grades cannot be the only method used to assess student achievement and understanding, test grades comprise a majority of a student's grade at both Mapledale and Stanhope (70% of their overall grade). Since universities make acceptance decisions based largely on grade point average, do students

who attend a school that use traditional grading or standards-based grading have an advantage?

Summary

After my experiences as a student and the opportunities I've had to assess students using both traditional and standards-based scales, I will be investigating these two grading practices. I will use both past and current studies to guide my analysis, and I will perform action research experiments in my own classroom. Additionally, I will survey students, parents, teachers, and administrators in both systems to gain insight on the perceived advantages and disadvantages of different stakeholders.

In this chapter, I have given the background of my interest in standards-based grading, and in Chapter Two I will review the current literature in this area. Chapter Three will focus on the methods I will utilize in my research, including participants and data collection methods. The results of my study including a summary of my findings will be outlined and discussed in Chapter Four, and in Chapter Five I will share the conclusions that I have reached through the research and analysis of my study.

CHAPTER TWO LITERATURE REVIEW

Introduction

The goal of this Capstone is to identify the impact of a standards-based grading system on student grades and achievement. As a teacher, I have used both the traditional and standards-based scales to assess students, but I have done so in two different schools. I intend to do a direct comparison between the systems so that I can answer the question, "What can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?"

Before comparing the two types of grading systems, traditional and standards-based, it is important to understand the history and key components of each. The concepts of both traditional and standards-based grading discussed in this chapter have been prepared through analyzing and reviewing literature written by educational researchers, educators, districts, and institutes of higher learning. This chapter has been separated into topics directly related to the two grading systems and the adoption of a standards-based grading policy as follows: history of grading, the traditional grading scale, standards-based grading, standards-based grading protocol adoption and adaptation, and the effect of standards-based grading on student engagement and motivation.

History of Grading

While the majority of schools and districts use percentage-grading scales that translate to letter grades A through F, this has not always been the case in education. Prior to 1850, grading did not exist. Children attended school and were grouped in one-room schoolhouses where all students, regardless of age and background, received instruction from the same teacher. Reporting on student progress occurred mostly orally between teachers and parents during home-visits by the teachers (Guskey, 2013, p. 68). As laws were passed in states requiring compulsory public education, school enrollment increased forcing public schools to adjust their approach to education. This ballooning enrollment led to age-grouping and the beginning of formal grading. At this point, teachers communicated students' development to their parents via narratives that described students' progress towards specific skills (O'Connor, 2010, p. 38). Between 1870 and 1910, the number of public high schools greatly increased from 500 to over 10,000, which led to the need for further organization of students. The formation of subject-area grouping occurred in high schools due to high demands on teachers (Guskey, 2013, p. 68). Now communicating performance became the responsibility of the subject area teachers. By 1910 schools transitioned to percentage grades, especially in high schools, because of the need for a more efficient grading practice (O'Connor, 2010, p. 38). However, a study conducted by Daniel Starch and Edward Charles Elliot in 1912 and 1913 brought into question the percentage method of grading, as they found that a wide range of grading practices existed among teachers (as cited in Guskey, 2013, p. 69). In response, the A through F grading scale was developed to reduce the number of categories through which to report student grades. This method became the mainstay in education until the 1990s when online grade books became

popular (Guskey, 2013, p.69). As schools adopted online grade books, grading systems transitioned back into percentage reporting because computers utilize number systems and are unable to calculate letter-based grades. However, schools could not abandon the A through F scale because Grade Point Average (GPA) is calculated using a 4.0 scale that directly ties to A, B, C, D, and F grades. Since colleges use GPA as a major factor in their admittance criteria, percentage grades had to directly align to these A, B, C, D, and F markers so that GPA could be calculated and compared across school districts (Guskey, 2013, p. 72).

The Traditional Grading Scale

The resurrection of percentage-based grading has begun anew the discussions surrounding grading reliability. Though percentages are easy to calculate, the alignment to A through F grades does not equally disperse the percentages over the entire 100-point scale. Generally, 90-100% is considered an A, 80-89% a B, 70-79% a C, 60-69% a D, and anything below 60% is failing. These means that out of 100 opportunities, students have 40 opportunities to receive a passing grade (60-100%) and 60 to receive a failing grade (below 60%). Looking at this another way, we have more failing grades than the number of passing grades combined (Hooper & Cowell, 2014, p. 61). Additionally, each band of mastery A through D is limited to only 10 percentage points, while all levels of failure fall into the 60 percentage points at the bottom of the scale. For example if a student receives a 91%, he or she has earned an A, whereas if a student earns an 89%, only two percentage points lower, his or her grade is a B. This 2% change has a large effect on a pupil's GPA because an A translates to a 4.0, and a B translates to a 3.0, but the difference in the level of student understanding is not that large of a gap. However, if

a student receives a 50% and another student receives a 20%, we as teachers do not specify a difference in that child's grade. Both are considered failing grades and result in a 0.0 in the student's GPA. The problem with this is that a test grade of 20% does far more damage to the overall grade than does a 50% and is much harder from which to recover mathematically, since when averaged with a higher grade, the 20% has a much more negative affect than the 50%.

Another criticism of percentage grading is that it is inconsistent with most colleges and universities, where integer grading is the method of choice. Colleges and universities have found that it is much easier and more representative of a student's aptitude to convert from a grading scale using 0, 1, 2, 3, and 4 (integers) to a GPA because it does not require converting percentages (Guskey, 2013, p. 72). This direct translation from integers to grade point average also removes some of the subjectivity found in percentage grading, specifically by removing much of the use of partial credit

Finally, in order to objectively grade students, Malehorn (1994) says that it is important that the students are aware of exactly what information we as teachers want them to master (p. 324). We provide learning targets and goals, teach the skills to reach those targets and goals, provide feedback through formative assessments throughout a unit, scaffold learning, review subject matter, and then test them on this knowledge. However, when we grade the tests, we are not doing so in a way that explicitly pairs their learning with the goals we have provided. As we grade assessments and award partial credit, it is impossible to do so in a fair and equitable manner, even if we have determined in advance how many points each part of a question encompasses (Stiggins, 2007, p.63). This leads to grading an assessment and trying to fairly delineate

between a student losing a point or a half of a point and then determining their grade by finding the percent correct they have achieved. Often, teachers grade a portion of an assessment and realize they need to adjust the way that they are awarding points, thus requiring the grading of a single student's assessment multiple times. When partial credit is awarded, we are not really communicating a student's mastery of a concept, but instead counting the number of correctly answered questions and questions that are partially correct and assigning a percentage based on numerical calculation. Educators in all subject areas agree that percent correct does not always indicate percent mastery (Guskey, 2013, p. 69). For example, a student could earn a 60%, which is a D. The student may not understand most of the concepts, but he or she may have adequately mastered some of the concepts, though the percent grade does not reflect that mastery. Further, a student may not know what concepts they understand or do not understand, as their percentage grade communicates only that they understood part of the unit taught.

If we recognize that percentage grading is so inherently flawed, why do we continue to use it as our most popular grading method as an educational community? According to Shippy, Washer, and Perrin (2013) our attitudes on grading are directly related to and based on how we were graded as students (p. 14). Since most educators are products of the percentage-based system, that is what is most familiar and thus most-utilized. Guskey, Jung, and Swan (2011) also purport that there is a lack of training for teachers on grading and reporting, so they have little choice but to replicate their experiences (p. 53). Additionally, the purpose for grading has not been clearly defined in the educational community (Hooper & Cowell, 2014, p. 60). In the traditional

model, grades are used to communicate learning, communicate effort, sort students, motivate students, punish students, or any combination of these. The traditional grading system is often used to create hierarchies within a classroom or school instead of for its original purpose, which was to emphasize student proficiency and prepare students for higher-level classes (Spencer, 2012, p. 4). This results in many students either being confused as to how their work is being measured or in pupils learning how to "play school" and achieve high grades but not actually master the material (Scriffiny, 2008, p.71). When we as teachers ask students to reach a bar that is different from district to district, school to school, grade level to grade level, and even teacher to teacher, we are not communicating their progress in the subject area as much as we are collecting evidence from both their performance and behavior in the classroom and averaging them together. Perhaps Guskey (2013) said it best when stating, "Percentage grading systems that attempt to identify 100 distinct levels of performance distort the precision, objectivity, and reliability of grades" (p. 72).

Standards-based Grading

Muñoz and Guskey (2015) state that "Grading represents teachers' evaluations - formative or summative - of students' performance" (p. 64). Standards-based grading differs from traditional grading in that it focuses solely on academic performance. In the traditional grading method used by the majority of schools across the country, a grade encompasses more than just student mastery of the subject matter. Homework, behavior, attendance, notebooks, group work, etc. are often factored into a student's final grade at the end of a quarter, semester, or year. The main problem with this is students do not necessarily know how each of these areas impacts their grade, making it difficult for

them to determine how to appropriately prioritize their time and learning (Kohn, 1994, p. 41). Standards-based grading looks to remove much of this ambiguity for both students and teachers.

Standards-based grading saw its inception more than thirty years ago, and as early as 1973 Paul Hirst discussed how vocational education assessed students based on "new" competencies exclusively (as cited in Shippy, Washer, & Perrin, 2013, p. 14). This idea drastically differs with the more traditional grading methods, because it focuses only on what we as educators want students to know and master. The key to assessing students only based on their subject knowledge is identifying learning targets and measuring students against how well they can perform to those targets. Elementary schools were the first to more widely implement standards-based grading by limiting grade book entries to standards only (Townsley, 2013, p. 70) because grading at an elementary level discusses student progress instead of focusing on an A through F grade in a certain subject area. This is much more difficult at the secondary level because percentage grades are utilized by most online grade books. When a teacher has more than 100 students, he or she often depends upon the online grade book to calculate student grades to keep the process as streamlined as possible.

Standards-based grading encourages teachers to design units with the end in mind (Shippy, Washer, & Perrin, 2013, p. 15). Learning targets within a unit are determined based on the content standards and then those learning targets are communicated to students at the beginning of the unit. Throughout the unit, students measure their own progress, similar to a workplace evaluation (Shippy, Washer, & Perrin, 2013, p. 16). Additionally, students have multiple opportunities to show mastery. In the current

practices in most districts, one unit test exists that measures a student's understanding at a particular point in time. However, it takes some students longer to learn certain concepts than others. One of the premises of standards-based grading is that a student can attain mastery at any point, and that does not have any impact on their final grade (Guskey, 2001, p. 21). If a student needs mastery of the current information to move on, especially in classes like math where much of the curriculum builds on itself, the student relearns the material he or she did not master alongside the new material. Ultimately, by the end of the year, the student's grade will show how much and which of the standards were mastered.

The main points of a standards-based grading protocol are: a grade is aligned to a single standard (Hooper & Cowell, 2014; Muñoz & Guskey, 2015; Shippy, Washer, & Perrin, 2013); grades are communicated on a scale of zero to four (Hooper & Cowell, 2014; Marzano & Heflebower, 2011); only the most recent scores are considered in a final grade (Hooper & Cowell, 2014; Marzano & Heflebower, 2011; Scriffiny, 2008); rubrics are provided to students and parents so that students can track their own progress (Scriffiny, 2008; Spencer, 2012; Townsley, 2013); and mastery is measured at many different points depending on the amount of time it takes a student to understand a concept (Hooper & Cowell, 2014; Shippy, Washer, & Perrin, 2013; Spencer, 2012). Many experts have detailed each of these points, and I will summarize their work below.

The first ideal in standards-based grading, and in fact its cornerstone, is that a grade is aligned to a single standard. This does not mean that for each individual benchmark, there is a grade associated. Currently, there are an average of 30 standards in

each state across 14 subject areas with over 3500 benchmarks (Scherer, 2001, p. 15). For a teacher to attempt to grade each of these individual benchmarks would be virtually impossible. Standards-based grading requires educators and schools to drastically reduce the number of standards in a content area. This also clarifies standards for students so they know the marks that they must hit. There can be subsections within these standards, but the measurement should be based on only three to five main standards per subject area (Muñoz & Guskey, 2015, p. 66). These three to five standards should be overarching standards for the subject area, and the current list of benchmarks would fall within those standards. For example, in a high school pre-calculus class the three main standards may be Algebraic Manipulation, Graphing, and Mathematical Communication. The specific benchmarks and skills for a pre-calculus class like graphing sine and cosine functions or proving trigonometric identities would fall under the Graphing standard and Algebraic Manipulation categories respectfully.

The second idea in standards-based grading is that grades are communicated on a scale of zero to four, or integer scale, where zero is matched with limited to no understanding of a concept, and four means that a student has fully mastered a concept (Hooper & Cowell, 2014; Marzano & Heflebower, 2011). Part of the reason for using this integer scale is clarity of the different achievement levels. When using traditional methods of grading, it is difficult to establish much difference between two percentage points on a 100-point scale. However, when only given five options for communicating understanding, expectations are much clearer for students because they can more easily see the difference between each achievement level. Additionally, the achievement levels are detailed so students have an understanding of exactly where they fall and what it

takes to move from one level to another. Further, these levels zero to four align exactly to letter grades A, B, C, D, and F providing a much more straightforward understanding of what each letter grade means (Guskey, 2013, p. 72).

Only considering the most recent score in a final grade is the third main idea in standards-based grading. This is a major departure from the method familiar to most teachers, students, and parents. Because most of us are products of a traditional system, we are accustomed to learning specific standards within a certain unit and then being tested on those standards at the end of the unit (Shippy, Washer, & Perrin, 2013, p. 14). This grade goes into the grade book and is averaged into a student's final grade. Though relearning and retesting is being used in some systems that use a traditional grading scale, this idea originally came from standards-based grading. Standards-based grading allows for relearning and retesting opportunities, and only the most recent grade is counted in a student's final mark. This provides latitude for knowledge to continue to grow throughout the course of a quarter, semester, or year, and the grade that is communicated is the student's level of mastery, no matter when that mastery occurred. While some schools that utilize the traditional grading system do employ relearning and retakes, the grading of these is still very subjective allowing for "partial credit." Additionally, in the systems that use this type of grading there is generally only one opportunity for relearning and retesting, which does not meet the criteria of continuous learning (Iamarino, 2014, p. 4).

A fourth major component of standards-based grading is a rubric for students and parents to track progress through a unit. In the traditional grading protocol, we often see that students do not have understanding of what they are learning day to day and how that

fits into the larger picture or the unit (Shippy, Washer, & Perrin, 2013, p. 15). In standards-based grading, students are provided with rubrics at the beginning of a unit that outline the learning targets and skills that a student must know to master that unit. Then, students are asked to measure their progress along the way using the same zero to four scale, thus taking responsibility for their own learning. The self-awareness students develop through assessing their own progress is an important life skill that students gain outside of the content knowledge that they are expected to master. Multiple formative assessments that are broken down by benchmark occur through the unit, and by providing constant feedback, we as teachers are communicating to students how they are doing on each benchmark, which they can then compare to their self-evaluation of their skills. All of this develops motivation and personal ownership of learning in students, thus enhancing their ability to prepare for summative assessments, or the accumulation of the unit's benchmarks, and ultimately self-assess their own level of mastery within the unit (Shippy, Washer, & Perrin, 2013, p. 15).

A final piece that is important in standards-based grading is the idea that students are assessed multiple times if needed to attain mastery. In my opinion, this is the most important of the components of standards-based grading. Human beings are always learning, so to ask a student to perform at a specific point in time and take that score as their only demonstration of understanding is in direct opposition to the goal of education. It penalizes students that struggle for needing extra time to learn the subject matter (Scriffiny, 2008, p. 73). The standards-based grading protocol provides retesting and relearning opportunities for students as many times as necessary to attain and demonstrate mastery. Some question why a student would work to learn something the

first time if he or she can relearn and retest as many times as necessary to receive a grade with which they are satisfied. The key is making certain assignments or learning tasks required prior to the first assessment and then requiring specific relearning tasks prior to receiving the opportunity to retest. In other words, a student is only eligible to retest if he or she has completed all of the necessary assignments prior to the first assessment and then also completes the relearning assignments prior to retesting (O'Connor, 2007, p. 239). While it is a cornerstone to this protocol, it is, in my opinion, probably also one of the more difficult aspects to achieve. Schools and school schedules are not currently constructed in such a way that allows a continuous learning process. A school year has a specific start and end date, and each year is broken into semesters and quarters, at which point we report "final" grades for our students. We then promote students at the end of the year to the next grade level if they have achieved a satisfactory grade in each subject area, regardless of their understanding and mastery level. While the school year, quarter, and semester cannot be adjusted at a school or classroom level, we can provide multiple opportunities for students to show mastery, and in order to implement standards-based grading, providing these opportunities is necessary.

With all of these requirements, implementing standards-based grading is a tall order. However, if we are asking students to be lifelong learners, it is also incumbent upon us as educators to evolve in our practices.

Standards-based Grading Protocol Adoption and Adaptation

The research that exists regarding standards-based grading is focused mostly on the importance of changing our grading practices and how a standards-based grading protocol improves student engagement and achievement. While much of the literature

discusses the need to create standards and grade based on those standards, allow students to relearn and retest, remove all non-summative grades from grade books, and engage students in the grading process throughout, there is not much information on how to create and grade assessments that do just this. Additionally, to effectively implement a standards-based grading practice into a school or district, teachers and administrators must together develop the process so that it is common across the school or district so that all teachers, administrators, parents, and students are speaking the same language. For the purposes of this section, I will discuss the process that a district or school might use to develop a standards-based grading protocol.

According to Guskey (2007) in the book *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning,* grades are a reflection of only the summative assessments within a class (p. 25). Formative assessments, homework, or any other learning tool used between summative assessments should only provide feedback to students and should not be included in the cumulative grade. Knowing that the summative assessment is the only grade that is communicated on a report card, it is imperative that designing a unit begins with the knowledge of what will be assessed at the end. As stated by Guskey (2007) some critics of standards-based grading have suggested that designing a unit based on the summative assessment is considered "teaching to the test" (p. 18) However, students are much more likely to be successful reaching mastery during a summative assessment if their learning goals for the unit are clear from the beginning (Guskey, 2001, p. 23).

Ainsworth (2007) suggests the following steps in creating a standards-based grading system. First, begin with "Power Standards," which are described as the most

important, high-impact standards in the subject area curriculum (p. 86). The number of Power Standards should be limited to three to five for clarity. Next, group key concepts and benchmarks under each of the Power Standards. Since subject areas are broken down into many various benchmarks, sometimes well into the hundreds, it is important to group them within the Power Standards. This ensures that each benchmark will be covered but that students and teachers know that these benchmarks are just a piece of one of the three to five main standards for a class. The third step Ainsworth suggests is to create common formative assessments. Many researchers in the area of standards-based grading argue that frequent formative assessments with specific feedback lead to higher overall achievement on summative assessments, and thus higher overall grades. The formative assessments need to be aligned to the benchmarks and then to the Power Standards so that as students receive feedback, they can assess their own progress toward mastery of a benchmark or standard. Next, the group must design a scale that clearly communicates student progress toward the standards. Most experts in the area of standards-based grading agree that a zero to four scale (or integer scale) works best. O'Connor (2007) further specified a possible beginning point for this scale, which are suggested using the following levels: Level 4, Advanced; Level 3, Proficient; Level 2, Approaching; Level 1, Beginning; and Level 0, Not Yet Attempted (p. 133). Marzano suggests breaking these levels down one step further to half points (as cited in Reeves, 2007, p. 113). While Marzano's levels are not technically considered integers (whole numbers), the suggested levels are much more detailed, and thus make the grading process more straightforward. Marzano's suggested scale is written as follows:

Score 0.0 - Even with help, the student demonstrates no understanding or skill.

Score 0.5 - With help, the student demonstrates a partial understanding of some of the simpler details and processes, but not of the more complex ideas and processes.

Score 1.0 - With help, the student demonstrates a partial understanding of some of the simpler details and processes and some of the more complex ideas and processes.

Score 1.5 - The student demonstrates partial knowledge of the simpler details and processes, but there are major errors or omissions regarding the more complex ideas and processes.

Score 2.0 - There are no major errors or omissions regarding the simpler details and processes, but there are major errors or omissions regarding the more complex ideas and processes.

Score 2.5 - There are no major errors or omissions regarding the simpler details and processes, and partial knowledge of the more complex ideas and processes.

Score 3.0 - There are no major errors or omissions regarding any of the information and/or processes (simple or complex) that were explicitly taught.

Score 3.5 - In addition to Score 3.0 performance, the student demonstrates partial success at inferences and applications that go beyond what was taught.

Score 4.0 - In addition to Score 3.0 performance, the student demonstrates in-depth inferences and applications that go beyond what was taught.

Both scales are 4.0 scales, but it is clear from the details listed above that Marzano's is much more detailed than O'Connor's. Each expert or researcher in the field has his or her own suggestion on how to design a rubric, so it is incumbent upon the district or

school to discuss and determine what best meets its needs. This includes determining the scale that best fits the school or district's grading practices and ensuring that as formative and summative assessments are created, they fit the grading rubric. While there is disagreement on the actual scale and rubric when defining achievement levels, all researchers all agree that specific learning objectives must be aligned to the scale and rubric so that students and teachers know at the beginning of a unit how mastery will be measured.

Once a grading scale is developed, benchmarks aligned, and summative and formative assessments created, teachers construct units based upon the benchmarks and assessments. A key part of the curriculum should be a list of goals for students to allow them to self-assess throughout the unit. By making clear what will be assessed at the outset, students become more engaged in the learning process because they have an understanding of the learning goals. Throughout the unit, the formative assessments indicate to students their level of understanding and show them where they can make improvements prior to the summative assessment (Ainsworth, 2007, p. 96). Additionally, the formative assessments should have similar rigor to the summative assessment. This way, as students receive feedback from formative assessments, they can build on the knowledge they gain through the feedback in order to better prepare for the summative assessments by comparing their previously completed work to the benchmarks communicated at the beginning of the unit.

While there are not many "steps" in the process of creating a standards-based grading protocol, each must be discussed and agreed upon by the teachers and administrators in a school or district and then further specified within content areas

(Hooper & Cowell, 2014, p.75). This is the process that is most difficult, as it requires agreement among all impacted parties, which is often very difficult to achieve.

The Effect of Standards-based Grading on Student Engagement and Motivation

Student engagement and motivation have become increasingly important in the discussion of student achievement. Emphasis no longer rests solely on grades and it is incumbent upon teachers to ensure that grades reflect student understanding and engagement. The growing popularity of standards-based grading has been a catalyst for including the discussion of student engagement in grading, as a cornerstone in the standards-based grading protocol is student self-assessment.

Imagine a high-achieving high school student the night before a summative assessment preparing and studying his or her homework and notes. Based on the assigned problems and what this student perceived as emphasized in class, he or she has prioritized study topics. Fast-forward to the next day when the student sits down to take the test and realizes that the topics he or she thought were important are only a small portion of the assessment. The student takes the test and receives a low grade that does not at all reflect his or her preparation. No matter how motivated and engaged this student is, he or she is likely to lose some of that motivation if it feels as though the hard work is not matched with the achieved grade. This is a common story in many schools that use the traditional, percentage-based grading system. Eventually, students lose motivation and trust in their teachers (Guskey, 2007, p. 17) and eventually see school and the learning process as a game instead of as a lifetime pursuit (Scriffiny, 2008, p. 71).

By reimagining our curriculum, we can have a strong positive effect on student motivation and engagement. According to Shippy, Washer, and Perrin (2013) grading

according to level of mastery of standards increases motivation. Designing activities and assessments by concepts improves student awareness and understanding of mastery, thus improving student ownership (p. 14). Scriffiny (2008) suggests that students are more likely to complete homework in a standards-based grading system, even if they do not receive any points toward their final grade for completing the homework, because they can see how the assignments are connected to assessments and learning targets instead of just looking at them as a way to gain points in a class (p. 72). This clarity builds student self-confidence because students know what is expected of them. In fact, when students at Arizona State University were surveyed regarding whether they preferred the traditional points-based grading method or a standards-based grading method, 86% preferred standards-based grading because they felt more ownership of their learning (Lord, 2016, p. 34).

Many researchers have suggested methods of improving student motivation using a standards-based grading practice. Malehorn (1994) suggests several ideas that make students part of the teaching and learning process by purporting that students should help choose materials and topics in each unit and participate in the evaluation process (p. 324). Asking students to be a part of the decision-making and assessment processes makes them stakeholders in their own success. When a person becomes a stakeholder, his or her feeling of ownership increases, which usually leads to improved engagement and mastery. Further, when students take part in their own assessment, they become more self-aware, which prepares them for life after school and to be more self-sufficient (Scriffiny, 2008, p. 73). In most career fields in which students will enter, self-assessment and self-awareness are expectations in the workplace. The standards-based

grading method mirrors the practice in the workplace of setting goals, and then assessing oneself based on goal achievement at the end of a specified time period. If we as educators make this process part of the daily school routine, students will obtain significant practice in self-assessing, so they will be much better prepared to enter the workforce and continue this process.

Iamarino (2014) further argues that increased student engagement increases subject-matter comprehension and understanding (p. 5). The argument that in addition to including student voice in the learning process, increasing formative assessments also improves motivation and comprehension is a valid one. Formative assessments provide an opportunity for a student to receive feedback. The more formative assessments in which a student participates, the more feedback the student receives. The specific, actionable feedback helps a pupil to focus on areas of need and what is required to improve in those areas. Then, students and teachers work together to establish connections instead of a teacher simply assigning points for completing a task. When a teacher administers a summative assessment, students see this as a chance to show what they know and how their knowledge and skills have improved, which holds the students accountable to their own progress. Thus, the grade earned on an assessment communicates comprehension instead of simply the ability for a pupil to complete assignments. When an overall grade is based only on summative assessments, it tells the story of a student's progress over the course of the grading period and his or her comprehension and understanding of the subject matter. This grading practice emphasizes the idea of continuous improvement, an idea utilized in all post-education endeavors. A points-based system runs counter to the model of continuous

improvement. Because a student in a traditional grading model earns points for turning in homework, completing quizzes, and doing well on tests, he or she can quickly adopt the mindset that says, "If I get enough points, I can move on," instead of one that sees value in the process of continuous learning (Iamarino, 2014, p. 4). Further, by allowing students to relearn and retake assessments, continuous improvement becomes a natural part of the process. A student can see in which area he or she has attained mastery and can then work on improving in the areas in which he or she did not perform well during the first assessment opportunity. This also mirrors what students will see when they enter the workplace. During each evaluation opportunity in a given career, part of the process is discussing areas of strength and areas in need for improvement. As one improves aspects of performance in one area, it is then incumbent upon him or her to find another area in which improvement can be made. By introducing students to this continuous improvement process during their school years, we are helping to make the transition to their careers and the workplace happen that much more smoothly.

By including students in the processes of designing units, choosing learning topics, evaluating their progress, and assessing, the level of understanding and engagement both increase. As students are more engaged, they become more motivated to improve, and eventually they achieve at higher levels. Scherer (2001) states, "Standards hold the greatest hope for significantly improving student achievement." (p. 14). In addition, students become more prepared to tackle their careers and the working world because they have been introduced and become familiar and adept at the self-assessment and continuous improvement processes.

Conclusion

This section reviewed literature that discussed the history of grading and the key aspects of the traditional grading methodology and the key components of a standards-based grading system. Also discussed were the adoption and adaptation of a standards-based grading system and the effect of standards-based grading on student engagement and motivation. The research in this Capstone will utilize the literature reviewed to create a standards-based grading rubric. This rubric will be used to assess student knowledge and achievement and compare the level of achievement with the level of achievement measured by a traditional grading scale. This data will be analyzed to answer the question, "What can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?"

Moving to a standards-based grading protocol requires changing more than just the method of grading assessments. While many of the important features are already in place in the district in which I will conduct my research, it is important to note that this may not be the case in all districts wishing to adopt a standards-based grading practice. The research is not intended to be used as a guide for implementing a standards-based grading practice in a school or district but instead as a comparison of the data gathered by each system of grading. By comparing assessment data side-by-side for each system, I endeavor to provide more information for educators, schools, or districts considering moving from a traditional grading to a standards-based grading protocol.

CHAPTER THREE METHODS

Introduction

There are many components of a Standards-based grading system: reconfiguring curriculum in such a way that it is organized within three to five benchmarks in a subject, writing assessments aligned to the benchmarks (Muñoz & Guskey, 2015, p. 66), and creating a rubric to effectively grade the assessments based on an integer scale of zero to four (Hooper & Cowell, 2014; Marzano & Heflebower, 2011). Students take an active role in the assessment process, which increases student motivation and achievement (Shippy, Washer, & Perrin, 2013, p. 15). Finally, an important aspect of the Standards-based grading philosophy is the opportunity for students to relearn and participate in reassessment in an effort to show mastery of each benchmark in the course content (Hooper & Cowell, 2014; Shippy, Washer, & Perrin, 2013; Spencer, 2012). Because most schools use the traditional percentage grading method, including the one in which my research will take place, I would like to see the effect of grading using a Standards-based rubric on overall student achievement. The question I have developed to investigate this idea is, "What can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?"

In this chapter, I will discuss the research paradigm I chose, setting for my research, methods used, tools, and data analysis methods. The setting section will also describe the participants and Human Subjects review information. The tools and data

analysis methods will give an overview of how I will be comparing traditional grading to Standards-based grading and what conclusions can be drawn from the data I collected.

Research Paradigm

I used the quantitative research paradigm described by Creswell in his book *Research Design: Quantitative, Qualitative, and Mixed Methods Approaches* (2014). I utilized unit assessments to collect data in the form of test averages on the unit assessments as well as overall percentage grade. Because I used an approach that collects data in numerical form and analyzes and compares this data, the quantitative research paradigm was the best choice for my Capstone. Additionally, my results were statistically analyzed for a quantitative approach to my question.

I used a quasi-experiment for my Capstone. I was a member of two Professional Learning Communities (PLCs), one for each of the two subject areas I taught (Geometry and Precalculus). Within these PLCs the use of common assessments was required, which meant that I could not use a different assessment as a dependent variable. Instead, I created a rubric to grade my assessments using a standards-based grading protocol, which served as the dependent variable. Because I only have access to my students' tests, the first semester exams of one of my Geometry sections and one of my Precalculus sections were used. The population could not be randomly assigned and was instead an "intact group" that was easily accessible to me as the researcher, thus necessitating the use of a quasi-experiment (Creswell, 2014, p. 219).

Due to the nature of my research, the choice of the quantitative research paradigm was that which best fits my needs. Using this methodology allowed to collect and analyze numerical data from my population within the classroom setting.

Setting and Participants

The school where my research took place is a suburban high school, Mapledale High School (pseudonym) that services students in grades nine through twelve. There are 1,807 students that attend this high school with the demographic makeup of 0.7% Native American, 16.3% Asian, 2.8% Hispanic, 4.6% Black, and 75.6% White. The school also has a free and reduced-price lunch population of 11.9%, and 7.5% of the school's population receives special education services.

I compared traditional and standards-based grading methods for assessments in both my Geometry and Precalculus classes. Both classes have students from all grade levels, nine through twelve. The Precalculus class is also a concurrent enrollment class with a local community college. This class is not required in the math curriculum, so all students in the class have taken it by choice. Geometry is a required math class for all students and must be successfully completed to be eligible for graduation. It is the second in the series of three math classes required by the State of Minnesota for graduation. The majority of students that take the class are freshmen and sophomores, but there are some upperclassmen in the class as well. Some students bypass this class by either completing an accelerated program, but otherwise all students of all ability levels are required to take this course.

Due to the use of previously completed assessments, informed consent is not required from students and parents. I have included in Appendix A a copy of the letter to the school administration outlining my research plan and methods.

Research Methods and Data Collection and Analysis

I was interested in identifying what, if any, difference exists in the measurement of student achievement when a traditional grading method is used versus when a standards-based rubric is used to grade end-of-unit assessments. To do this, I graded my students' assessments twice, once using the traditional percentage correct method, and once using a standards-based rubric.

My first step was to create a rubric using Marzano's grading methods, a 0-4 scale, as seen below.

Standards-based Rubric

Standards-based Grade	Correct Questions	% in Gradebook
4	75% R2 & 75% R3 & R4	100%
3	75% R2 & 75% R3	95%
2.5	75% R2 & 50% R3 or 75% R3	85%
2	75% R2 Or 50% R2 & 50% R3	75%
1.75	50% R3	65%
1.5	50% R2 Or 25% R2 & 25% R3	60%
1	Anything below	50%

This rubric is created using questions called "R2, R3, and R4." This "R" designation is the level of question: an R2 problem is a question that, if answered correctly, shows a basic understanding of a concept; an R3 problem is a question that requires one or two additional application steps to solve, which shows a more in-depth understanding of a

concept; an R4 problem is one that is two to three levels above the current class and shows a student's ability to synthesize and combine multiple concepts to solve a difficult problem. The "percent in gradebook" column is used to allow a direct comparison between traditional grading scores and standards-based scores. The standards-based rubric does not have any scores lower than 50% in an effort to ensure that all grade levels (A, B, C, D, and F) have equal weighting in the grading scale and to avoid having the number of failing grades be larger than the number of passing grades, which is seen in the traditional method.

Next, each assessment question was assigned an "R" designation depending upon the level of question. The assessments are currently written with problems that have different levels of difficulty, but in order to utilize a rubric, I needed to assign levels to each question to match those on the rubric. Additionally, the assessments are already broken into benchmarks, so they were easily translated into standards-based grading because each question only addresses a single benchmark.

After the completion of semester one, I gathered the tests from one section of my Geometry classes and one section of my Precalculus classes that had been graded using the traditional percentage correct method. Then, I went through the tests again and used the rubric to assign a standards-based grade that I then transferred to the associated percentage grade as seen in the rubric. By comparing these two grades, I was able to see if class averages are higher, lower, or the same with the two methods. Another factor that I will be comparing is the effect that standards-based grading has on students' grades in Precalculus compared to those in Geometry. Because of the setup of the math curriculum, one of the classes (Geometry) is required for graduation, while the other

(Precalculus) is taken by student choice, specifically by those who hope to master higher levels of mathematics than are required by the State of Minnesota for graduation. The difference in these two types of classes also causes a difference in the student population within the classroom. I analyzed whether the difference in course also affects overall student achievement when comparing traditional and standards-based grading methods.

Summary

My Capstone focuses on what one might learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method. I graded tests given in both Geometry and Precalculus classes using two different protocols: traditional percentage correct grading and a standards-based grading rubric. I then compared the test averages between methods as well as across types of classes to see the effect on the measurement of student achievement. In the next Chapter, I will discuss the results of my research, including the averages of each grading method and the comparison across class types. Additionally, I will suggest ideas for moving forward with the implementation of a standards-based grading system in a school or district. I will include my recommendations for assessment rubrics and other considerations for Standards-based Grading transition and utilization.

CHAPTER FOUR RESULTS

Introduction and Review of Research Question

The purpose of this study is to examine the differences when grading unit tests using a traditional grading scale versus a standards-based grading method. This chapter will provide an overview as to how the test scores were collected. It will also analyze the data to determine the differences that exist and their impact on overall student grade and overall unit exam average.

Chapter Three explained the methods used to research the question, "What can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?" The study utilized a quasi-experimental design to grade tests twice from two groups of students' (one section of Geometry students and one section of Precalculus students), first using the traditional grading scale and then using the standards-based grading method. The data was then analyzed to identify themes that existed on each unit test as well as on students' overall grades in the class, only taking these unit tests into consideration. I will start by discussing each class's results individually, and then I will look at a comparison between the two classes based on those results. For the purposes of this study, I have defined a grade difference as a change between full letter grades (from a B to an A for example) and did not look at the

difference of percentages within each of those letter ranges (A+ to A or A to A- for example).

Geometry

The Geometry class was made up of 32 students between grades nine and twelve. I graded five unit exams from first semester using both a traditional scale and a standards-based method to find if any difference existed in overall grade for students and in overall average grade for each individual test. Of the 32 students, 16 showed a letter grade increase when graded using the standards-based method versus the traditional method. Of the sixteen grades, one was only a 2% change (from an 89% up to a 91%), so this grade may have stayed the same when factors other than tests were averaged in to the student's overall grade at the end of the semester. The other 15 showed at least a 4% increase in grade, and the largest increase was 19%. Of particular note, with the traditional scale, five students would have failed the semester of the Geometry class. However, when graded using the standards-based method these students would have passed, likely due to the 50% floor in the standards-based method.

One of the 32 students' grades decreased when using the standards-based method in comparison to the traditional scale. However, this student's percentage landed very close to the grade letter boundary in both the traditional scale (70.635%) and the standards-based method (69%), so it is difficult to say whether the grading method would have had an affect on the overall grade the student received for the semester.

The remaining 14 students achieved the same letter grade using both grading methods. The largest change in percentage for these students was approximately 4%. Had the students with the 3-4% differences been closer to the grade letter boundaries,

their grades may have shown an overall change, but that was not the case for these students.

Overall, of the 32 students studied in this Geometry class, 50% (16 students) showed a grade increase, 46.9% (15 students) of students' grades remained the same, and 3.1% (one student) showed a decrease in grade when comparing the standards-based method to the traditional scale.

When I looked at the averages for the Geometry tests, I determined that I received better information looking at the percentages than at the letter grades. When I initially looked at the letter grades, four of the five showed an increase of one letter grade using standards-based method. Upon closer inspection, I noticed that all five average percentages were higher using the standards-based method than using the traditional scale. The smallest increase was about 2%, and the largest increase was about 7%. I believe this is also due to a minimum grade of 50% in the standards-based method.

Precalculus

The Precalculus class was made up of 24 students between grades nine and twelve. I graded six unit exams from first semester using both a traditional scale and a standards-based method to find if any difference existed in overall grade for students and in overall average for each individual test. Of the 24 students, 10 showed a letter grade increase when grading using the standards-based method in comparison to grading using the traditional grading scale. Of the ten grades, one was only a 0.1% change (from an 89.9% up to a 90%), and two students had a difference of 1%, so these grades may have stayed the same when factors other than tests were averaged in to the students' overall grades at the end of the semester. The other 7 showed at least a 4% increase in grade,

and the largest increase was 7%. In this class, all students would have passed the class using either grading method.

One of the 24 students' grades decreased when using the standards-based method in comparison to the traditional scale. This student's grade in the traditional scale was 80.5%, and with the standards-based method was 77.5% or a difference of 3%. This difference is likely due to my designation of question levels when grading using the standards-based method as well as the lack of partial credit in the standards-based scale.

The remaining 13 students achieved the same letter grade using both grading methods. The largest change in percentage for these students was approximately 5%. Like with the Geometry students, had the Precalculus students with the 3-5% differences been closer to the grade letter boundaries, their grades may have shown an overall change, but that was not the case for these students.

Overall, of the 24 students studied in this Precalculus class, 41.67% (10 students) saw a grade increase, 54.17% (13 students) of students' grades remained the same, and 4.16% (or one student) showed a decrease in grade when using the standards-based method versus a traditional grading scale.

As with the Geometry tests, I looked at the averages of each individual test based on percentage. Of the six unit tests I analyzed, five showed an increase in percentage when using the standards-based method compared to the traditional scale, and one showed a decrease. The increases were between approximately 3% and 6.5%, and the decrease was approximately 2%.

Geometry and Precalculus Compared

Looking at the comparison between the two classes actually gave more insight than looking at each class individually. Specifically, the percentage of student grades that increased was larger in the Geometry class (50%) than in the Precalculus class (41.67%), meaning that the change to standards-based grading affected more students positively in the Geometry class than it did in the Precalculus class. Further, there was a larger proportion of students that had a grade C or lower that increased their grade with the standards-based method. In Precalculus, 70% of the students that would have received a grade increase had a C or lower increased their grade when using the standards-based method, and in Geometry that percentage was 75%. Of the 14 students in Geometry whose grades stayed the same, only one student had a grade of C or lower, and in Precalculus, five of the 13 students had a grade of C or lower (approximately 38%). Looking at this data, it seems that the standards-based method helps students with lower grades improve their grade more than it helps students with higher grades. This is likely due to the lowest grade in the standards-based scale being 50% and is also related to the number of students failing Geometry varying so drastically between the traditional scale (5 students) and the standards-based method (0 students).

Conclusion

In this chapter, I have analyzed the data from grading tests for both a Geometry class and a Precalculus class in two different ways: once with the traditional scale and once using the standards-based methods. In both classes, approximately half of the students saw a grade increase using the standards-based method. This was especially true of students with a letter grade of C or lower. In the next chapter, I will discuss my

conclusions and detail further topics of research and studies to more specifically determine the differences between the two grading methods.

CHAPTER FIVE CONCLUSIONS

Introduction

The results from the study presented in Chapter Four aimed to answer the question: "What can one learn by measuring assessments in two ways: traditional grading scale versus standards-based grading method?" The study investigated differences between using the two grading protocols – traditional grading scale and standards-based grading method – and also compared the use of these methods between a Geometry class and a Precalculus class. I specifically addressed individual student grade changes and overall test percentage changes. Chapter Five will further investigate the results of the study and discuss limitations or problems encountered. I will include ideas for future research and reflect on what I learned during the Capstone process.

I obtained the study results by using unit tests from one Geometry class and one Precalculus class. I graded the tests using a traditional grading scale. I then assigned levels (R2, R3, R4) to each of the questions and graded the tests again using the standards-based grading method and rubric outlined in Chapter Three. I then compared and analyzed the data gathered from these two methods in the Geometry class and the Precalculus class and then compared the results between the two classes.

Summary of Findings

The primary goal of this study was to determine what differences one might find between grading unit tests using a traditional grading scale versus a standards-based method. In the Geometry class, approximately 50% of the 32 students showed an increase in letter grade when using the standards-based method over the traditional grading scale, and only one student showed a decrease. In the case of the decrease the difference in percentage was very small, so the grade may not have actually been different when other class work was averaged in for the semester. The remaining students' letter grades stayed the same with each grading method used. In all five of the unit tests, there was an increase in overall test average when using the standards-based method.

In the Precalculus class of 24 students, ten of them (approximately 42%) showed a grade letter increase using the standards-based method in comparison with the traditional grading scale. As with the Geometry class, one student showed a grade decrease, and the remaining 13 students had the same letter grade using either method. Five of the six unit tests showed an increase in test average when using the standards-based method.

When comparing the Geometry and Precalculus classes, approximately the same proportion of students saw a grade increase using the standards-based method. Upon further examination, the majority of students (70% for Precalculus, 75% for Geometry) with increased grades had a letter grade of C or lower using the traditional grading scale. Similarly, looking at the students whose grades remained the same, the majority of students (62% for Precalculus, 100% for Geometry) had a grade of B or higher.

From the data obtained and analyzed, I can conclude that for approximately half of students, using the standards-based method of grading would improve their grades, with the remainder of students' scores remaining the same, independent of method used. I can also conclude that the students who would be at the greatest advantage with the adoption of the standards-based rubric I used in my study would be those with letter grades of C or lower. This is likely due to the fact that 50% is the lowest grade a student can earn on any exam or assignment. This indicates that students who may have received very low grades (below 50%) using the traditional method were at a distinct advantage when graded using the standards-based method.

Implications

One of the fundamental criticisms of the traditional grading scale is that for each letter grade A through D (100% to 60%) there are only 10 opportunities out of 100 (or 10%) to receive a specific letter grade. However, there are 59 chances out of 100 to fail (receive lower than a 60%). Additionally, when averaging scores lower than 50% with scores above 59%, the lower the score, the more detrimental to the student's overall grade. With the standards-based rubric that I researched and used, an adjustment is made so that there is an equal percentage that falls in the category of "failing" as there is in each of the other letter grade categories. This led to several students in my study passing the class who would have failed using the traditional grading method.

Because the chance of failing when the standards-based method is used decreases so drastically, this grading method may have a significant effect on student motivation. For example, if a student does receive a failing grade on an exam, knowing that his or her grade would not be decreased so dramatically might encourage that student to keep trying

to study and improve his or her grade with each subsequent test. However, if a student receives a very low grade using the traditional method, anything below 50% for example, the student may realize that the grade will be much harder to improve because it is averaged with the other tests at the end of the semester, and thus he or she may "give up" and not put in as much effort to achieve a grade improvement.

Limitations

If I were to perform this study again, there are several changes I would implement. First, I was not able to use several of the unit tests from the Geometry class and the Precalculus class due to length. Because of the method that I chose for my standards-based grading protocol, a test had to have a minimum of eight questions to adequately align with the rubric. There were three tests between the two classes that only had seven questions or less, and thus could not be used in the study. Another change that I would make would be to adjust tests so that they would be in multiples of four questions prior to grading. Again, due to the standards-based grading protocols, all of which use a scale of 0-4 in some way, tests that did not have a multiple of four had to be reevaluated prior to grading them using the standards-based method. For example, if there were two questions that assessed the same skill, I removed one of those questions to achieve a multiple of four. Further, questions that have multiple parts do not lend themselves to a direct conversion to a standards-based grading method. On this type of question, it was necessary to assign each part of the multi-part question its own "R" designation to fairly grade students who may have made a mistake in an earlier part of the question, but got an answer that made sense at the end. Ideally, if one were to reproduce this study, it would

be wise to make changes to the tests in advance for easier comparison of grading methods.

Possible Future Studies

In the future, I would like to continue to research the difference between a traditional grading scale and a standards-based grading scale. The ideal method for this study would be a controlled experiment. This may require the approval of the school or district administration, as it would require authoring and giving two different tests to the same type of class (meaning two Precalculus classes or two Geometry classes). Specifically, it would require that one test be authored and graded with the idea of using the traditional grading scale and partial credit in mind. However, it would need to be written alongside the test that would be graded using the standards-based protocol so that the tests would be assessing the same skills. Then, one class would receive the traditional test, and the other class would receive the standards-based test. The comparison of the results between the two would more precisely show differences because the tests would be written specifically for the grading scale that would be used to assess them. Additionally, it would allow the comparison of data on individual questions if desired. In essence, the class taking the traditional unit test becomes the "control group" and the class taking the standards-based unit test becomes the "experimental group."

Another possibility for future study would be to look at student motivation. It would be necessary to develop a method to gather student opinions utilizing both grading methods. Again, I would develop tests for two different classes, one using traditional grading and the other using standards-based grading and ask for student feedback. I would be particularly interested in two areas. First, I would be curious to know how

students felt about each grading method, what questions they had, and what they would change about that particular grading method. I also would be interested to do a student-engagement survey to determine student motivation before any of the exams and then after each exam to see what differences were noticed as the semester continued.

Reflection

When I began this Capstone process, I was convinced that I would unequivocally determine a "best" grading method. As I researched and learned more about how the traditional grading scale was developed and read more about standards-based grading, I was convinced that at the end of the Capstone I would determine that standards-based grading was the "better" protocol and would have data and research to further demonstrate this belief. However, this is not what happened. If anything, I became more curious and interested in the difference between the two and would like to continue finding ways to research the two protocols to come up with a grading method that best reflects student knowledge.

The difficulty with any grading protocol is that we as educators are trying to measure student knowledge in a concrete, numerical way. The problem with this is that we each bring our own knowledge and biases into grading. When assigning partial credit, each teacher may approach this differently, even if the grading outline is discussed and agreed upon before the test. Because of this, I would naturally tend to think that the standards-based grading practice that I used is more appropriate as it does not use partial credit. Where I struggled with this, however, is what if a student makes a small mistake, like forgets a negative sign, which drastically changes his or her overall answer? If we are using the standards-based method, that question is wrong, even with just a small

mistake. With the traditional method we give partial credit to the student. Why not just use a traditional method then? Well, what if a student receives a 25% on a test? That student's grades must be much higher on subsequent tests in order to hope to pass the class. With the standards-based method, this is taken into account because there is a 50% floor. This is why this process has been so eye-opening for me. I figured I would have evidence to show that one method is better than the other. However, what I found is that I am left with many more questions than answers. My guess is that there is a balance between the two protocols, but I am not sure what that balance is, and it would definitely require more study.

What I have been able to conclude from my Capstone process is that I will be sharing my experiences with my students in the coming years. I will not specifically be explaining my data or which research method I used or what was included in my research process. Instead I have a renewed excitement in life-long learning that I will hope to demonstrate to my students. I want to encourage students to continue questioning to continue learning. This is, after all, what I most hope they will gain from their schooling, and I hope my personal growth from this process will help me to encourage the same in my students.

REFERENCES

Ainsworth, L. (2007). Common formative assessments: The centerpiece of an integrated standards-based assessment system. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 79-102). Bloomington, IN: Solution Tree.

Andrews, L. (2007). Grades, scores, and honors: A numbers game? *Journal Of The National Collegiate Honors Council*, 23-30.

Ash, K. (2012). Lessons from higher education. *Digital Directions*, 26.Shippy, N. N.,

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Davies, A. (2007). Involving students in the classroom assessment process. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 31-58). Bloomington, IN: Solution Tree.

Deddeh, H., Main, E., & Fulkerson, S. R. (2010). Eight steps to meaningful grading. *Phi Delta Kappan*, *91*(7), 53-58.

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership*, *59*(1), 20-27.

Guskey, T. R. (2004). 0 Alternatives. *Principal Leadership: High School Edition*, *5*(2), 49-53.

Guskey, T. R. (2007). Using assessments to improve teaching and learning. In D. Reeves

(Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 15-30). Bloomington, IN: Solution Tree.

Guskey, T. R., Jung, L. A., & Swan, G. M. (2011). Grades that mean something. *Phi Delta Kappan*, *93*(2), 52-57.

Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership*, *71*(1), 68-72.

Hardré, P. H. (2014). Checked your bias lately? Reasons and strategies for rural teachers to self-assess for grading bias. *Rural Educator*, *35*(2), 17-23.

Hooper, J., & Cowell, R. (2014). Standards-based grading: History adjusted true score. *Educational Assessment*, *19*(1), 58-76.

Iamarino, D. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education, 17*(2), 1-10.

Johnson, B. G., & Beck, H. P. (1988). Strict and lenient grading scales: How do they affect the performance of college students with high and low SAT scores?. *Teaching Of Psychology*, *15,* 127-131.

Kohn, A. (1994). Grading: The issue is not how but why. *Educational Leadership*, *52,* 38-41.

Lord, M. (2016). Made to measure. *ASEE Prism*, *25*(6), 32-35.

Malehorn, H. L. (1994). Ten measures better than grading. *Clearing House*, *67,* 323-324.

Marzano, R. J. (2007).Designing a comprehensive approach to classroom assessment. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 103-126). Bloomington, IN: Solution Tree.

Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know.

Educational Leadership, *69*(3), 34-39.

Marzano, R. J. (2012). An easier way to score tests. *Educational Leadership*, *69*(6), 82-83.

McIntosh-Burke, H. (2013). Four grades and no GPA. *Journal Of College Admission*, (219), 5.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan*, *96*(7), 64-68.

O'Connor, K. (2007). The last frontier: Tackling the grading dilemma. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 127-146). Bloomington, IN: Solution Tree.

Pollio, M., & Hochbein, C. (2015). The association between standards-based grading and standardized test scores in a high school reform model. *Teachers College Record, 117*(11), 1-28.

Rauschenberg, S. S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives*, *22*(92), 1-38.

Sawyer, R. R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement In Education*, *26*(2), 89-112.

Scherer, M. (2001). How and why standards can improve student achievement: a conversation with Robert J. Marzano. *Educational Leadership*, *59*(1), 14-18.

Scriffiny, P. L. (2008). Seven reasons for standards-based grading. *Educational Leadership*, *66*(2), 70-74.

Spencer, K. (2012). Standards-based grading. *Education Digest*, *78*(3), 4-10.

Stiggins, R. (2007). Assessment for learning: An essential foundation of productive instruction. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 59-78). Bloomington, IN: Solution Tree.

Townsley, M. M. (2013). Redesigning grading—Districtwide. *Educational Leadership*, *71*(4), 68-71.

Washer, B. A., & Perrin, B. (2013). Teaching with the end in mind: The role of standards-based grading. *Journal Of Family & Consumer Sciences*, *105*(2), 14-16.

APPENDIX A

Letter of Intent to Mounds View High School

December 2, 2016

Dear Mounds View High School Administration,

I am currently a mathematics teacher at this school and a graduate student working on an advanced degree in teaching at Hamline University, St. Paul, Minnesota. As part of my graduate work, I plan to conduct research using tests from Semester 1.The purpose of this letter is to give you written record of approval. This research is public scholarship, and as such, the abstract and final product will be catalogued in Hamline's **Bush Library Digital Commons**, a searchable electronic repository, and it also may be published and used in other ways.

I plan to study the use of Standards-based grading to grade past assessments from September 2016 – December 2016. This will involve grading student work using both the traditional grading method and a standards-based rubric. Students have already received grades for this class, so they will not be involved in any way.

I have received approval for my study from the School of Education at Hamline University. The capstone will be catalogued in Hamline's **Bush Library Digital Commons**, a searchable electronic repository. My results might also be included in an article for publication in a journal or in a report at a professional conference.

I have received a verbal agreement from the Principal of Mounds View High School, Dr. Jeffrey Ridlehoover. Please keep this as a written record of my research.

Sincerely,
Lynn S. Gay
1900 Lake Valentine Road, Arden Hills, MN 55112
651-621-7178
lynn.gay@moundsviewschools.org

APPENDIX B

Geometry Data

Geometry Test Data – Traditional Grading Method

Test 5 - Traditional	Test 4 - Traditional	Test 3 - Tradtional	Test 2 - Traditional	Test 1 - Traditional	PERCENT	LETTER GRADE
82.143%	75.000%	83.333%	65.517%	66.667%	74.532%	C
46.429%	40.000%	57.143%	41.379%	30.000%	42.990%	I
96.429%	90.000%	92.857%	74.138%	96.667%	90.018%	A
52.381%	32.500%	61.905%	63.793%	63.333%	54.782%	I
97.619%	72.500%	100.000%	81.034%	96.667%	89.564%	B
42.857%	62.500%	66.667%	39.655%	58.333%	54.002%	I
96.429%	75.000%	95.238%	84.483%	106.667%	91.563%	A
97.619%	87.500%	97.619%	63.793%	75.000%	84.306%	B
100.000%	90.000%	100.000%	86.207%	100.000%	95.241%	A
61.905%	72.500%	42.857%	67.241%	55.000%	59.901%	D
95.238%	92.500%	100.000%	82.759%	50.000%	84.099%	B
96.429%	80.000%	95.238%	87.931%	93.333%	90.586%	A
96.429%	100.000%	92.857%	94.828%	93.333%	95.489%	A
54.762%	75.000%	85.714%	43.103%	60.000%	63.716%	D
73.810%	65.000%	61.905%	75.862%	76.667%	70.649%	C
100.000%	92.500%	100.000%	84.483%	98.333%	95.063%	A
78.571%	92.500%	100.000%	72.414%	96.667%	88.030%	B
61.905%	87.500%	90.476%	86.207%	86.667%	82.551%	B
100.000%	100.000%	100.000%	93.103%	106.667%	99.954%	A
85.714%	20.000%	64.286%	77.586%	71.667%	63.851%	D
55.952%	40.000%	45.238%	56.897%	55.000%	50.617%	I
96.429%	70.000%	90.476%	68.966%	95.000%	84.174%	B
89.286%	92.500%	80.952%	60.345%	83.333%	81.283%	B
100.000%	87.500%	100.000%	94.828%	101.667%	96.799%	A
63.095%	70.000%	92.857%	70.690%	76.667%	74.662%	C
91.667%	85.000%	100.000%	75.862%	91.667%	88.839%	B
54.762%	75.000%	57.143%	60.345%	61.667%	61.783%	D
58.333%	87.500%	59.524%	25.862%	53.333%	56.911%	I
69.048%	55.000%	71.429%	81.034%	76.667%	70.635%	C
95.238%	95.000%	100.000%	81.034%	103.333%	94.921%	A
54.762%	55.000%	90.476%	68.966%	48.333%	63.507%	D
91.667%	87.500%	97.619%	65.517%	80.000%	84.461%	
79.278%	**75.078%**	**83.557%**	**71.121%**	**78.385%**		

Geometry Test Data – Standard-based Grading Method

Test 5 - Standard	Test 4 - Standard	Test 3 - Standard	Test 2 - Standard	Test 1 - Standard	PERCENT	LETTER GRADE
75.000%	85.000%	95.000%	60.000%	95.000%	82.000%	B
60.000%	75.000%	50.000%	60.000%	60.000%	61.000%	D
95.000%	95.000%	100.000%	95.000%	95.000%	96.000%	A
60.000%	50.000%	85.000%	75.000%	65.000%	67.000%	D
100.000%	75.000%	100.000%	85.000%	95.000%	91.000%	A
50.000%	75.000%	95.000%	60.000%	60.000%	68.000%	D
100.000%	75.000%	100.000%	85.000%	105.000%	93.000%	A
95.000%	85.000%	100.000%	65.000%	85.000%	86.000%	B
100.000%	95.000%	100.000%	85.000%	100.000%	96.000%	A
60.000%	85.000%	85.000%	75.000%	75.000%	76.000%	C
95.000%	95.000%	100.000%	85.000%	65.000%	88.000%	B
95.000%	85.000%	95.000%	95.000%	95.000%	93.000%	A
95.000%	100.000%	95.000%	95.000%	95.000%	96.000%	A
50.000%	95.000%	85.000%	60.000%	100.000%	78.000%	C
75.000%	65.000%	75.000%	75.000%	75.000%	73.000%	C
100.000%	95.000%	100.000%	85.000%	95.000%	95.000%	A
85.000%	95.000%	100.000%	85.000%	95.000%	92.000%	A
85.000%	95.000%	100.000%	65.000%	85.000%	86.000%	B
100.000%	100.000%	100.000%	95.000%	105.000%	100.000%	A
95.000%	50.000%	85.000%	75.000%	85.000%	78.000%	C
60.000%	60.000%	85.000%	60.000%	75.000%	68.000%	D
100.000%	85.000%	95.000%	75.000%	95.000%	90.000%	A
95.000%	95.000%	60.000%	75.000%	85.000%	82.000%	B
100.000%	95.000%	100.000%	100.000%	100.000%	99.000%	A
60.000%	75.000%	95.000%	85.000%	95.000%	82.000%	B
95.000%	85.000%	100.000%	95.000%	95.000%	94.000%	A
60.000%	85.000%	75.000%	75.000%	75.000%	74.000%	C
65.000%	95.000%	85.000%	50.000%	60.000%	71.000%	C
50.000%	60.000%	75.000%	75.000%	85.000%	69.000%	D
95.000%	100.000%	100.000%	85.000%	100.000%	96.000%	A
60.000%	75.000%	85.000%	85.000%	60.000%	73.000%	C
95.000%	75.000%	100.000%	75.000%	95.000%	88.000%	B

Geometry Letter Grade Comparison – Traditional vs. Standards-based Methods

TRADITIONAL	STANDARD
C	B
I	D
A	A
I	D
B	A
I	D
A	A
B	B
A	A
D	C
B	B
A	A
A	A
D	C
C	C
A	A
B	A
B	B
A	A
D	C
I	D
B	A
B	B
A	A
C	B
B	A
D	C
I	C
C	D
A	A
D	C
B	B

Indicates Grade Increase

Indicates Grade Decrease

Indicates No Grade Change

APPENDIX C

Precalculus Data

Precalculus Test Data – Traditional Grading Method

Test 6 - Traditional	Test 5 - Traditional	Test 4 - Traditional	Test 3 - Traditional	Test 2 - Traditional	Test 1 - Traditional	PERCENT	LETTER GRADE
70.370%	82.609%	93.750%	73.684%	79.630%	76.000%	79.340%	C
74.074%	100.000%	100.000%	84.211%	90.741%	90.000%	89.838%	B
44.444%	100.000%	76.563%	63.158%	75.926%	86.000%	74.348%	C
68.519%	91.304%	92.188%	57.895%	87.037%	78.000%	79.157%	C
57.407%	86.957%	89.063%	78.947%	85.185%	72.000%	78.260%	C
70.370%	100.000%	67.188%	71.053%	48.148%	64.000%	70.126%	C
55.556%	67.391%	90.625%	78.947%	83.333%	78.000%	75.642%	C
70.370%	95.652%	75.000%	86.842%	92.593%	92.000%	85.410%	B
62.963%	78.261%	92.188%	84.211%	90.741%	88.000%	82.727%	B
79.630%	86.957%	96.875%	97.368%	92.593%	96.000%	91.570%	A
72.222%	78.261%	81.250%	76.316%	66.667%	64.000%	73.119%	C
61.111%	100.000%	96.875%	81.579%	92.593%	86.000%	86.360%	B
92.593%	100.000%	100.000%	100.000%	100.000%	100.000%	98.765%	A
51.852%	73.913%	65.625%	84.211%	55.556%	84.000%	69.193%	D
72.222%	43.478%	85.938%	86.842%	87.037%	72.000%	74.586%	C
70.370%	86.957%	82.813%	86.842%	74.074%	82.000%	80.509%	B
62.963%	100.000%	89.063%	97.368%	98.148%	92.000%	89.924%	B
64.815%	67.391%	60.938%	76.316%	70.370%	88.000%	71.305%	C
25.926%	82.609%	82.813%	63.158%	61.111%	80.000%	65.936%	D
75.926%	82.609%	79.688%	100.000%	70.370%	86.000%	82.432%	B
79.630%	100.000%	96.875%	100.000%	94.444%	84.000%	92.492%	A
87.037%	100.000%	100.000%	100.000%	90.741%	88.000%	94.296%	A
42.593%	78.261%	81.250%	89.474%	85.185%	94.000%	78.460%	C
57.407%	52.174%	87.500%	89.474%	85.185%	84.000%	75.957%	C
65.432%	**84.783%**	**86.003%**	**83.662%**	**81.559%**	**83.500%**		

Precalculus Test Data – Standards-based Grading Method

Test 6 - Standard	Test 5 - Standard	Test 4 - Standard	Test 3 - Standard	Test 2 - Standard	Test 1 - Standard	PERCENT	LETTER GRADE
85.000%	85.000%	95.000%	85.000%	85.000%	85.000%	86.667%	B
60.000%	100.000%	100.000%	95.000%	95.000%	95.000%	90.833%	A
50.000%	100.000%	75.000%	85.000%	85.000%	100.000%	82.500%	B
60.000%	95.000%	95.000%	60.000%	85.000%	85.000%	80.000%	B
75.000%	85.000%	75.000%	75.000%	85.000%	75.000%	78.333%	C
85.000%	100.000%	75.000%	85.000%	65.000%	85.000%	82.500%	B
60.000%	75.000%	85.000%	65.000%	85.000%	95.000%	77.500%	C
75.000%	95.000%	85.000%	85.000%	95.000%	95.000%	88.333%	B
60.000%	85.000%	85.000%	85.000%	95.000%	95.000%	84.167%	B
85.000%	95.000%	95.000%	100.000%	95.000%	100.000%	95.000%	A
60.000%	85.000%	75.000%	85.000%	75.000%	75.000%	75.833%	C
60.000%	100.000%	95.000%	95.000%	95.000%	95.000%	90.000%	A
95.000%	100.000%	100.000%	100.000%	100.000%	100.000%	99.167%	A
60.000%	85.000%	75.000%	95.000%	65.000%	85.000%	77.500%	C
60.000%	60.000%	85.000%	95.000%	85.000%	85.000%	78.333%	C
75.000%	85.000%	60.000%	85.000%	75.000%	85.000%	77.500%	C
60.000%	100.000%	85.000%	100.000%	100.000%	95.000%	90.000%	A
75.000%	75.000%	50.000%	85.000%	75.000%	75.000%	72.500%	C
50.000%	85.000%	75.000%	50.000%	65.000%	95.000%	70.000%	C
75.000%	95.000%	85.000%	100.000%	75.000%	95.000%	87.500%	B
75.000%	100.000%	95.000%	100.000%	95.000%	95.000%	93.333%	A
85.000%	100.000%	100.000%	100.000%	95.000%	85.000%	94.167%	A
50.000%	85.000%	85.000%	95.000%	95.000%	100.000%	85.000%	B
60.000%	50.000%	85.000%	95.000%	85.000%	85.000%	76.667%	C
68.125%	**88.333%**	**83.958%**	**87.500%**	**85.417%**	**90.000%**		

Precalculus Letter Grade Comparison – Traditional vs. Standards-based Methods

TRADITIONAL	STANDARD
C	B
B	A
C	B
C	B
C	C
C	B
C	C
B	B
B	B
A	A
C	C
B	A
A	A
D	C
C	C
B	C
B	A
C	C
D	C
B	B
A	A
A	A
C	B
C	C

Indicates Grade Increase

Indicates Grade Decrease

Indicates No Grade Change