6-1-2019

The Accuracies of Assessment-based Grading Compared with Total-points Grading Practices in Secondary Math Education

Emily Winter
Concordia University - Portland, ewinter@fairburyjeffs.org

6-2019

The Accuracies of Assessment-based Grading Compared with Total-points Grading Practices in Secondary Math Education

Emily Winter
Concordia University - Portland

Concordia University–Portland

College of Education

Doctorate of Education Program

WE, THE UNDERSIGNED MEMBERS OF THE DISSERTATION COMMITTEE

CERTIFY THAT WE HAVE READ AND APPROVE THE DISSERTATION OF

Emily Winter

CANDIDATE FOR THE DEGREE OF DOCTOR OF EDUCATION

Barbara Weschke, Ph.D., Faculty Chair Dissertation Committee

Julia Britt, Ed.D. Content Specialist

Tony Goss, Ph.D., Content Reader

The Accuracies of Assessment-Based Grading Compared With Total-points Grading Practices in Secondary Math Education

Emily Ann Winter

Concordia University–Portland

College of Education

Dissertation submitted to the Faculty of the College of Education

in partial fulfillment of the requirements for the degree of

Doctor of Education in

Higher Education

Barbara Weschke, Ph.D., Faculty Chair Dissertation Committee

Julia Britt, Ed.D., Content Specialist

Tony Goss, Ph.D., Content Reader

Concordia University–Portland

2019

Abstract

The purpose of this study was to evaluate the correlational relationship between classroom grades and standardized test scores for Grade 8 math students. Grading plays a central role in the field of education; this study was significant as it addressed a gap in grading research comparing two current grading methods, and provided an evidence-based discussion on grading practices. The classroom grades for 56 Grade 8 math students were calculated from the same classroom assignment and assessment scores using two grading strategies, a total-points method and an assessment-based grading method, with consideration of summative test retake scores. The assessment-based grading method explicitly employs tenets of Bloom's (1968) theory of learning for mastery. A correlational analysis found a positive relationship between all classroom grades from both total-points and assessment-based grading methods and standardized MAP math test scores. A Pearson correlation analysis determined the strongest statistically significant relationship was between total-points classroom grades and standardized MAP math test scores. Of the grading systems analyzed, the traditional total-points grades were the best representation of student learning when measured against standardized test scores. A summary of the implications from this study relates the results back to Bloom's theory of mastery-learning and provides recommendations on best-practices for classroom grading.

Keywords: grading practices, total-points grading, assessment-based grading, grading accuracies

Dedication

This dissertation is dedicated to my family, who instilled a passion for learning that inspired me to pursue this journey, and provided the encouragement, love and support to see me through.

Acknowledgements

I would like to express my gratitude and appreciation to my chair, Dr. Barbara Weschke. You were instrumental in helping me achieve my vision for my research, were responsive and quick with feedback, and pushed me to evolve my dissertation into its best version. I would also like to thank my committee members Dr. Julia Britt and Dr. Tony Goss. You may not have known it at the time, but Dr. Britt, you always seemed to offer words of encouragement exactly when I needed to hear them the most. Dr. Goss, I appreciate your shared passion for my topic. Your own experiences with district grading practice changes made your input invaluable during this process. I also could not have done this without the cooperation of the participating teacher. Not only did you grant me access to the data that made my research possible, you are one of the most impressive teachers I have ever had the pleasure to work with. You are the epitome of what we all strive to be. I would like to thank my long-time mentor, Dr. Larry Scharmann. I pray that you know the impact that you have had on my life, and that it was you that inspired the start of my doctoral journey. To my school lunch tribe, you are the network that I need to make it through each day. I am grateful for your relentless support, encouragement, friendship, and humor. You are each a blessing to me. To my daughters, Delilah, Mae and Pippa, you are the best thing I have ever done. I hope that I have made you proud. I would like to thank my husband, Jon, my biggest cheerleader and my rock. My sister and best friend, Amanda, who has always been fiercely protective and encouraging. I think that you are incredible. To my parents, both teachers and lifelong enthusiasts of learning, you have never doubted my ability to accomplish anything that I was determined to do. This is for you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Introduction to the Problem

Classroom teachers' grading has been a central issue in education, and the popularity and of different grading systems has waxed and waned over time (Brookhart et al., 2016). Grades are symbols assigned to student work or a composite value representing student knowledge (Brookhart etl al., 2016). Configuring and assigning grades serve a necessary role; however, there are frequently debates within districts, among teachers, and among researchers about how a grade should be determined and what it should represent (Hanover Research, 2011). Akins (2016) found that many principals want to make grading changes and administrators disagreed on perceptions of grading and of different grading systems. There are often inconsistencies in grade determinations among teachers.

Starch (1913) conducted a foundational grading study that found large variability between teacher-assigned grades and notable differences between standards of different schools. A student may receive two different letter grades from two different teachers for the same performance (Wormeli, 2009). School stakeholders question which instructor's grade is right. Educational researchers work to ascertain which grading practice is the most accurate representation of student learning. Teachers may wonder how grade values compare to a testing benchmark standard. These are questions faced by educators every day in the classroom setting; therefore, this study sought to provide evidence to help inform classroom grading decisions.

According to Brookhart (2011), a student's grade should represent mastery of corresponding content learning objectives. Academic grades are a vital part of the education system and play a central role in the educational experience of all students (Brookhart et al., 2016). School officials use grades to rank students, differentiate which students should be

awarded college acceptance, and document a satisfactory level of learning in classes (Brookhart et al., 2016). Similarly, grades are used to award course credits at the collegiate level and used to distinguish exemplary students (Schneider & Hutt, 2013). Academic grades have the potential to influence a student's future and to provide opportunities or possibly limitations on post-secondary opportunities. Grades predict important future educational consequences such as college success or dropping out of school (Brookhart et al., 2016). Because of their significance and potential consequences, grades should not be calculated with a casual approach or loose standards (Muñoz & Guskey, 2015). Instructors have a responsibility to students to accurately assign grades in a meaningful way.

This researcher will use this introductory chapter to provide an explanation of the historical trends in grading practices, the context of these grading concerns, and an introduction to the conceptual framework of mastery learning. Next, a statement of the problem and purpose of the study will set the stage for this study. The researcher has outlined research questions to represent the objective of this comparative grading analysis and an explanation of the significance and relevance will justify the importance of this study. The researcher provides definition of terms to define and clarify common grading terminology. Collectively, the researcher will use this chapter to set the stage for the proceeding literature review and study methodology.

Background, Context, History, and Conceptual Framework for the Problem

By the early 1900s, most schools and universities in the United States were using a 100-point grading system (Schinske & Tanner, 2014). The traditional total-points grading approach has the longest tradition in education. Citing a survey in 1998, O'Connor (2009) reported 91% of schools used a traditional grading system. The rise in popularity of the 100-point grading

system inspired educational research on the accuracies of grading, which questioned the consistencies in teacher-assigned grades (Schinske & Tanner, 2014). Traditional grading practices have remained unchanged for nearly a century and were instituted when only an advantaged few were expected or allowed to advance to post-secondary opportunities (Carifio & Carey, 2010). By the 1940s, the A–F scale system had emerged and gained a foothold (Schinske & Tanner, 2014). Letter grades were an efficient way to indicate a level of performance to students, and to serve as a tool for teachers to categorize students into ability groups (Vatterott, 2015). Letter grades and corresponding GPAs are still a common method used to rank students for college acceptance and scholarship awards (Guskey, 2014). Teachers and practitioners who view grades as a comparative tool used to rank students and determine placements on academic and vocational tracks may be more supportive of a traditional total-points grading practice (Vatterott, 2015).

Throughout the evolution of the current education system, a total-points grading system has been a popular choice and is still a common grading practice in classrooms today. Most students, teachers, and parents are familiar with a total-points grading approach. Therefore, it can be considered the default grading system (O'Connor, 2009). In a traditional total-points grading system, the greater value an assignment or assessment has, the more points the instructor assigns to that assignment. To determine a student's percentage grade for the grading period, the total points earned are divided by the total points possible, and the decimal is converted into a percentile value. Consequently, a summative assessment would naturally be worth more than an assignment when calculating the classroom grade; however, the importance and weight of any assignment depended on the points value assigned. It is typical that teachers may assign a larger point value to longer or more difficult assignments or tests.

For example, a homework assignment may be worth 10 points, but a project, paper or test could be worth 100 points. In calculating the academic grade, the test would count for more of the total points available. A natural weighting through point-assignments has been the predominant grading practice used by most instructors at all grade levels (O'Connor, 2009). O'Connor and Wormeli (2011) questioned how the total-points grade may be influenced by nonacademic factors. In a primary and secondary classroom, it was not unusual for teachers to include extraneous factors that influenced the academic grade (O'Connor & Wormeli, 2011). Some teachers give extra-credit opportunities for different tasks that are either alternative or extension content activities, or for trivial tasks such as a student bringing a box of tissues for the classroom. Other factors also may have had an impact on the grade tabulation.

Many teachers penalized late assignments from students with a reduction in points awarded (Wormeli, 2009). Students could also receive zeros if they were under an out-of-school suspension. Nonacademic factors such as behavior, attendance and effort influence the student grade in some grading models (Wormeli, 2009, & O'Connor, 2009). Advocates stated that nonacademic factors should not occlude the composition of an academic grade, a value that should solely represent student content knowledge and mastery of learning objectives (Marzano & Heflebower, 2011). Concerns with the inaccuracies in traditional grading practices (Merki & Holmeier, 2015) opened the door to alternative grading practices.

Bloom (1968) developed a theory on learning for mastery, later referred to as mastery learning, that explained how, given an elongated instructional window with remediation and repeated opportunities to demonstrate mastery, more students can be successful at a higher level of achievement. Bloom (1976) provided evidence and a theoretical framework that influenced a shift in grading practices to a more criterion-referenced approach (Schinske & Tanner, 2014).

Criterion-referenced grading systems evaluated students against learning objectives without a comparison to other students (Great Schools Partnership, 2014). Standards-based grading, a type of criterion-referenced grading, first gained popularity in 1983 and was born of an outcome-based reform movement (Stanford University, 2000). As Bloom's (1968, 1974, 1976, 1982, 1984) theory of mastery learning gained traction and support in the field of education, it inspired new practices with performance-based, outcome-emphasized strategies such as standards-based and assessment-based practices (Kalnin, 2014). The push for a more proficiency-based grading system came in 1998; the idea grew in popularity and evolved into the standards-based grading (Kalnin, 2014). By 2013, standards-based grading had grown exponentially (Kalnin, 2014).

Colleges in the United States reported that a student's completed coursework, test scores, class rank and grade point average, or GPA, are among the top determinants for college admissions, with student grades noted as the most significant factor (College Board, 2018). Standards-based grading uses a performance rubric and performance indicators instead of traditional percentages and letter grades. College officials have stated their support for standards-based instruction, but acknowledged the difficulties faced by college admissions staff and scholarship committees because of the absence of grade data traditionally submitted for each student, including GPA and class rank (Stefanowicz, 2016). Practitioners recognized the value of standards-based grading, but the grading method did not provide the student grading data that stakeholders were most familiar (Stefanowicz, 2016). A derivative of standards-based grading evolved as assessment-based grading which maintained the emphasis on summative performance but provided the numerical grade data, including percentile, corresponding letter grades, and GPA, which schools are more familiar (O'Connor, 2012). Kalnin (2014) explained that the emphasis on an outcome-based approach, including both standards-based and assessment-based

grading, was inspired by Bloom's (1968, 1974, 1976, 1984) work. This explanation provides a credible link that establishes a relationship between the instructional practices based on mastery learning that are used in both standards-based and assessment-based grading practices.

The newest of these grading practices is assessment-based grading. Assessment-based grading, as explained by O'Connor (2012), emerged and started to gain popularity at the secondary level around 2010. An assessment-based grading policy differed from a total-points approach because it had weighted categories with formative scores for homework, daily work and quizzes comprising no more than 20% of the classroom grade. Comparatively, summative assessment scores constituted at least the remaining 80% of the academic grade as a weighted category. Within each of the weighted categories, individual assignments can be weighted with total points. For example, a 100-point summative test would count for a greater portion than a 40-point summative assessment within the weighted category comprising at least 80% of the academic grade.

Proponents of the assessment-based grading dictate weighted categories of at least 80% summative assessment; however, the teacher or school may mandate an even greater weight in the assessment category (O'Connor, 2012). The second differentiating factor that distinguishes assessment-based grading from a traditional total points system is the option for students to retake a summative exam for a higher replacement grade (O'Connor, 2012). Some instructors in a total-points grading system offer retakes; however, the recorded test grade is an average of the two test grades or capped at a certain replacement score. In an assessment-based grading practice, retakes on summative assessments are encouraged and any higher score replaces and overrides the lower initial value.

While cooperative learning is encouraged as a formative activity, group work should not be recorded as a summative grade for students in assessment-based grading, as it cannot accurately evaluate the individual performance of a student (O'Connor & Wormeli, 2011). This stipulation is one of the ways that assessment-based grading aims to prevent nonacademic factors from influencing the academic grade, which was a concern with previous grading strategies (O'Connor & Wormeli, 2011). The primary differences in assessment-based grading compared with the traditional total-points system were the option to retake material for a higher grade and the greater emphasis and weight of summative assessment grades on the final academic grade calculation.

Conceptual Framework

Both assessment-based grading and standards-based grading policies implement strategies based on Bloom's (1968) theory of learning for mastery. Bloom's (1974) theory of mastery learning states that given supportive instructional practices, 90% of students can achieve mastery of the learning objectives. The differentiating instructional practices that support mastery learning begin after a traditional end-of-unit assessment. After an assessment, any students who demonstrate mastery on the first attempt are offered enrichment activities. Students that do not demonstrate mastery are to be provided with individualized reteaching and remediation, then offered an opportunity to retake the assessment to demonstrate gains in mastery. These steps lengthen the window for instruction and provide individualized correction. Given these supports, the theory of mastery learning states that a greater percentage of students will achieve proficiency of the content learning objectives, than with traditional instruction that does not offer extended opportunities to learn and demonstrate increased mastery.

Bloom's (1968, 1974, 1976, 1982, 1984) theory of mastery learning serves as the conceptual framework for these grading practices because both strategies include offering students repeated opportunities to demonstrate learning, which elongates the learning window. Both the assessment-based and standards-based grading strategies also provide interventions and supports to students who do not initially demonstrate proficiency for the learning objectives in support of Bloom's (1968) theory of learning for mastery. Test retake opportunities included in the assessment-based grading practice promote and facilitate content mastery, according to the theory by Bloom (1968). This link provides an explanation for the selection of mastery learning as the conceptual framework for this study. Kalnin (2014) explained how assessment-based grading and standards-based grading were both forms of proficiency-based grading, a phenomenon linked to mastery learning and outcome-based education.

The school district and state that provided the setting for this study will be referred to by the pseudonym Midwestern, to protect the confidentiality of participants. In the Midwestern state that provides the setting for this study, the math curriculum as outlined by the Midwestern State Math Standards, is cyclical; it scaffolds each year based on students' previous knowledge as is outlined in the progression of indicators in the state standards and curriculum guide (Midwestern Department of Education, 2018). It is important for students to demonstrate proficiency for grade-level specific learning objectives because future learning objectives are based on the presumed mastery of previous math content. To best achieve student mastery for the greatest percentage of students, a grade practice that is more supportive of mastery learning may be more appropriate. There is a substantial body of research supporting the effectiveness of a mastery learning approach, first proposed by Bloom (1968). In the upcoming literature review in Chapter 2, the researcher has included a synthesis of research on the effectiveness of mastery

learning instructional strategies. A mastery learning approach is often used as a mechanism to successfully implement a standards-based method. Less research was available on assessment-based grading, as it is a newer grading practice. This study presented an opportunity to add to the growing research repertoire on assessment-based grading by comparing the relationship between grades derived from assessment-based grading and total-points grading approaches.

Statement of the Problem

This researcher conducted an empirical review of the current research on grading practices and identified a deficit in research on the accuracies of assessment-based grading in comparison to traditional total-points grading. Assessment-based grading has increased in popularity since 2010 (O'Connor, 2012). There is currently a gap in research that compares assessment-based grades to standardized test scores; more research is needed to inform the recommendation for the promotion or discontinuation of assessment-based grading as a classroom best-practice. This study used a quantitative approach to identify if there is a relationship between the eighth-grade student grades derived from both assessment-based and total-points grading practices in comparison to standardized test scores.

Purpose of the Study

The purpose of this correlational research was to evaluate the accuracies of two major grading practices used in secondary schools. Assessment-based grading grew in popularity starting in the early 2000s (Kalnin, 2014). Many school districts, including the district that provided the setting for this study, are now requiring teachers to use assessment-based grading for learning practices as a school-wide grading policy (XXXXX & XXXXXXXX, 2017). If a school is going to require a teacher to use a mandated grading policy, it is most responsible to promote a research-supported grading practice. This study worked to discover the type and

strength of relationship between assessment-based grades and standardized test scores. It is important to evaluate the accuracy of this grading policy to ensure that grades, which can impact a student's educational journey, are an accurate representative of student learning, consistent, meaningful and supportive of learning (O'Connor, 2012). The results of this research study may be relevant and applicable to current topics in the field of education, should be considered by administrators when choosing a district-wide grading policy, and reflected on by classroom teachers when assigning student grades.

Research Questions

This study was guided by the central research question: Which grading method determines classroom grades for Grade 8 math students that most accurately align, with the smallest variance, to corresponding standardized math MAP test score values for Grade 8 math students? Under this central research question, the specific research questions addressed in the study include:

1. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students?
2. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students?
3. What was the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students?

4. Which grading method (traditional total-points grading and assessment-based grading) has resultant scores that more closely aligned to students' standardized test scores?

Rationale, Relevance, and Significance of the Study

The reason for this study was to specifically compare two grading methods currently used in schools. These results will add to the growing body of grading research. Grading and grades play such a large role in education that it is imperative that grades are accurate, consistent, and representative of student learning (Wormeli, 2009). This study is significant because it is difficult to find research available that directly compares assessment-based grading and total-points grading methods. The results of this study will be applicable to classroom teachers and stakeholders involved with grading decisions.

Rationale of the study. The topic of grading has always been a central conversation in the field of education, as is evident by the extensive body of research with grading themes (Brookhart et al., 2016). Academic grades are widely used for a variety of applications and comparisons; however, there is too large of a variance in the method of computation (Tinkelman, Venuti, & Schain, 2013). This lack of standardization in how grades are calculated causes inconsistencies in the accuracy and validity of the grades assigned. While most professionals agree what an academic grade should represent, all too often the academic grade is influenced by outside factors such as student behavior (Hanover Research, 2011). These inconsistencies may challenge the integrity of a grade and question its value.

Relevance of the study. If grading practices were standardized through the implementation of a single grading policy, the assigned grades would become more meaningful and accurate for comparisons (O'Connor, 2012). Student grades can provide feedback for

curriculum decisions and instructional effectiveness. Academic grades can be used to make recommendations for students for certain degree programs or career paths (Guskey, 2011). Students can use grades for self-evaluation and goal setting. Grades play a significant and central role in the education system, so it is imperative that the most effective grading practice is selected for implementation. This research may provide data to support which grading practice is indeed the best choice for classroom teachers to use.

Significance of the study. The significance and implications of students' grades have ensured that grading remained a central issue in the field of education (Brookhart et al., 2016). A number of researchers attempted to identify the most effective grading practices to assess student learning. Denson (2013), Fink (2015), Lee (2013), and Comes (2015) provided evidence that supports the usage of traditional total-points grading practices. Iamarino (2014), Brophy (2013), Knaack, Kreuz, and Zawlocki (2012), Hardegree (2012), Beatty (2013) and Timmons (2017) demonstrated growing evidence-based support for a standards-based grading approach. A standards-based approach provided an appropriate context for the discussion on assessment-based grading because it was similar in methods and goal outcomes.

The recent research from 2010 through 2018 on standards-based grading included in the following literature review in Chapter 2 demonstrate that research continues to support the effectiveness of a standards-based approach from a content-mastery perspective. Researcher data established the effectiveness of standards-based grading practices (Heflebower et al., 2014). Additionally, research also showed positive feedback on standards-based grading from both student and parent perspectives (Knaack, Kreuz & Zawlocki, 2012). It is relevant to include recent standards-based grading research studies in the following literature review in Chapter 2 because there were more literature resources available on this topic than for assessment-based

grading and the two approaches are comparable in methodology and theory. These studies, which the author will present in the literature review, demonstrate conflicting feedback regarding the effectiveness of each grading practice. These discrepancies in the literature highlighted the need for this research study, which conducted a direct correlational analysis of different grading practices on a single study population.

Definition of Terms

Assessment-based grading: or "grading for learning" refers to a grading practice that calculates the student academic class grade based on a total points system within weighted categories with summative assessments comprising no less than 80% of the overall academic grade and the availability of retakes on summative assessments with any higher grade replacing the previous test score (O'Connor, 2012).

Criterion-referenced assessment: an evaluative description of learning objectives to be assessed without reference to the performance of others (Great Schools Partnership, 2014).

Formative assignment: deemed to be any assignment or activity that is elicits evidence of the current level of day-to-day understanding for students, information from formative assessments is often used to provide data to teachers to influence instructional decisions (Goodrich, 2012).

Grade: symbols assigned to student work or a composite value representing student knowledge (Brookhart et al., 2016). Grades may consist of a letter on the A-F scale, a percentile value, or label for a performance level (O'Connor, 2009).

Grading: the evaluation of student achievement on a larger scale, either for a single piece of work or for an entire course, subject, unit or module within a program and often serve as

material for grade determinations. Grading symbols may be letters A-F, or descriptive terms (Nitko & Brookhart, 2011).

Grading practice: a mathematical approach used to determine a student's classroom grade, often resulting in a percentile or letter grade; often includes educational philosophical implications related to test retake opportunities and types of activities to be scored and included in the grade. Examples include total-points grading and standards-based grading (O'Connor, 2009).

Measures of Academic Progress (MAP): national, norm-referenced, content-specific adaptive computer test to assess student's content grade-level knowledge (NWEA, 2018).

Outcome-based grading (OBG): a grading practice that evaluates student learning against a set of learning outcomes, usually performance-based and with expanded opportunities for students to demonstrate knowledge (Brandt, 2018).

Standards-based grading (SBG): the grading practice that continually evaluates student performance, assessing the level of student mastery on learning concepts. Student performance is documented using rubrics tied to specific learning objectives (Marzano, 2010).

Rasch UnIT (RIT): norm values for the NWEA MAP test (NWEA, 2018).

Score: also known as a raw score, is the letter or number given to any student test or performance that provides evidence of student achievement (O'Connor, 2009).

Summative assessments: classified as occurring at the end of instruction and practice as a final comprehensive of individual student learning of the learning objectives. Summative assessments may include practical, skill evaluations and performances in addition to traditional paper or online assessments but each student should be assessed on an individual basis (O'Connor, 2012).

Total-points grading system: term refers to a grading practice that calculates the student academic class grade based on a percentage of total points earned versus total points available without any weighted categories and with the assumption that there are not test retake opportunities for higher replacement scores (Tinkelman et al., 2013).

Assumptions, Delimitations, and Limitations

The results of this study relied on a few basic assumptions. It may be assumed that the classroom teacher was using authentic assessments to accurately evaluate student learning. It may be assumed that each student was given a fair opportunity to demonstrate his or her knowledge. It may be assumed that the classroom instruction was in support of the determined learning objectives; the validity of standardized test scores may also be assumed.

The delimitations for this study included the parameters for selecting student assessment scores and the time frame from which the data was accessed. Another boundary of the study included the selection of the study population, including which, if any, student scores were to be excluded from the data set and analysis. The study retrieved eighth-grade students' scores data from the 18-week spring semester of the 2017–2018 school year. The 2017–2018 fall semester was not selected as this researcher wanted to best assess the cumulative summative learning for students as reflected in both the academic grades and standardized test scores. This was best accomplished by using the second semester, end of the year, scores.

The limitations of this study were centered on identifying a standard benchmark to compare each of the two final grades to, one calculated using assessment-based and total-points grading practices. The NWEA Math MAP test was chosen as the standardized test in this study because of its sound reputation and categorical disaggregation of student data (NWEA, 2018). All standardized tests, however, are vulnerable to score inaccuracies due to outside factors.

There are potential issues, as with any standardized test, of students exerting minimal or no effort when taking the MAP math test. This would result in an inaccurate assessment of that student's content knowledge, which would consequently prove to be an erroneous benchmark with which to compare classroom grades. The same limitations would apply to any standardized test. The MAP math test is administered in the same classroom and environment as the instruction and classroom assessments occur, so there are no concerns with environmental influences that are not also present for the other data collection points.

Summary

There is a need for more research on providing evidence-based support for a grading practice that best fits the needs of teachers to most accurately represent and encourage student learning. There are positive character traits that teachers work to instill in students outside of their academic content. It is this researcher's perspective that, while these are decidedly important lessons for students too, they should not be reflected in, or influence the determination of, an academic grade. Some educational specialists believe that academic grade should solely be a representation of an individual student's mastery of learning objectives and a measure of content knowledge (O'Connor & Wormeli, 2011). If this serves as the purpose of a grade, it is important to solicit research on the assessment-based grading practice to see if it proves, through the quantitative analysis of data, that a grading policy that incorporates master learning principles is more accurate and meaningful for students.

For teachers who are required to summarize and communicate a student's academic performance through a single report card grade, either a percentile or letter grade assignment, there must be a consensus that the grade represents the student's most accurate academic achievement, and only reflects academic achievement (O'Connor, 2009). This study sought to

evaluate which grading system is more accurate, traditional total-points grading or assessment-based grading in comparison to a standardized test score benchmark. This was best accomplished through a correlational analysis. The results of this study may be valuable to classroom teachers and administrators looking to choose a grading system that is the most accurate representation of student content knowledge.

In upcoming chapters, this researcher will provide a summary of current grading research in Chapter 2. This literature review will include a focus on the most recent research regarding the validity of total-points grading, standards-based grading, assessment-based grading, and mastery learning. Next, Chapter 3 will explain the selection of a correlational study as the methodological design for this study in an effort to find evidence of a relationship between two grading practices and the standardized test scores of students. Chapter 4 will provide the data analysis and results that addressed the outlined research questions. Chapter 5 will present and discuss the findings of this study and includes a summary of the implications and recommendations from these results for further study.

Chapter 2: Literature Review

Introduction to the Literature Review

In this literature review, the researcher will present an appraisal and synthesis of the current literature most applicable to assessment-based grading and the research on standards-based grading as its outcome-driven precursor. Research on the effectiveness of instructional techniques associated with the theory of mastery learning will provide the lens to explain the practices and strategies that serve as the basis for both standards-based and assessment-based grading practices. The researcher will use the literature review to outline the criteria for standards-based grading as the primary conceptual framework for this research study. The researcher will also provide an explanation for the significant influence of the theory of mastery learning as it provided the evidence-based foundation for both standards-based and assessment-based grading practices. An outline and summary of current research studies and empirical reviews will be divided into categories that first address the studies on standards-based grading, followed the current research on assessment-based grading practices, traditional total-points grading practices, and lastly the research-driven support for mastery learning. This literature review will provide the context and background to serve as a setting for the interpretation and application of this research.

This research study sought to evaluate which grading practice, traditional total-points or assessment-based, is more accurate of student knowledge for eighth-grade math students, in comparison to corresponding state standardized math test benchmarks. This researcher's review demonstrated that more research has been dedicated to evaluating the standards-based grading practice, which cannot comprehensively be applied to the practice of assessment-based grading. While the two grading practices are similar, they are not homologous enough for the literature

regarding standards-based grading to be blindly extended and applied to assessment-based grading practices. This researcher's literature review identified a research gap in the accuracies of assessment-based grading. The synthesis of the literature review will provide insight on the themes identified throughout the literature review and justify the significance of this research topic.

Research search terms were centered on accuracies in classroom grading practices, accuracies of standards-based grading, accuracies of assessment-based grading, and correlations between summative grade outcomes and standardized testing scores. The researcher will use the literature review to provide the context of mastery-oriented, standards-based grading as the conceptual framework for this research. This research study presents an opportunity to increase and strengthen the current body of research on the accuracy assessment-based grading in direct comparison to traditional total-points grading as evaluated against a standardized test score. This comprehensive explanation of the topic of assessment-based grading and related grading practices will allow practitioners to confidently make instructional decisions based on the application of the research discussion and results.

Conceptual Framework

This research study was a grading comparison centered on the grading practices of total-points grading and assessment-based grading. The conceptual framework will explain the practice of assessment-based grading and its foundational theory, learning for mastery. The theoretical basis for standards-based grading and assessment-based grading, which stems from the extended window for instructional and assessment practices, is an implementation of the theory of mastery learning by Bloom (1968, 1974, 1976, 1982, 1984). In its original conception, the theory of mastery learning was defined in the context of cumulative student achievement

over specified content learning objectives (Block, Airasian, Bloom, & Carroll, 1971). Bloom (1968) asserted that, given the proper circumstances that extended the instructional period and extended opportunities for instructional gains, over 90% of students could learn the course content or learning objectives at a mastery level. This learning-for-mastery approach creates a feedback loop, increasing motivation for students to continue to strive to learn a greater proportion of the class content and struggling or reluctant learners are provided with additional supports. The mastery-learning approach (see Figure 1) extends the instructional time and opportunities for students to demonstrate knowledge (Bloom, 1968).

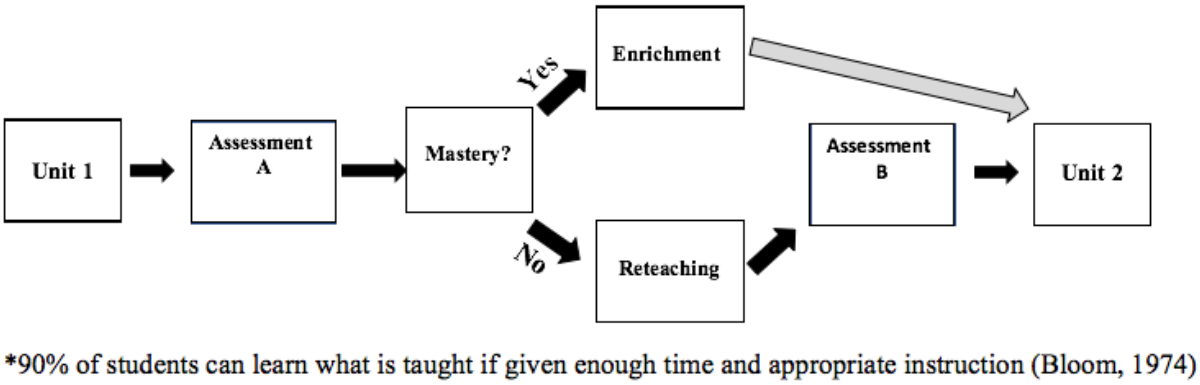

Figure 1. Adapted from Bloom's (1968) theory of mastery learning.

The incorporation of learning-for-mastery strategies in the assessment-based and standards-based grading practice justified a review of the research that provided evidence in support of learning-for-mastery techniques. The construct of mastery learning was an appropriate selection as the conceptual framework because it rationalized the inclusion of a test-retake practice in the assessment-based and standards-based grading policies. This process elongated the instructional window and provided reteaching opportunities. The steps of reteaching and reassessment in these grading practices are a direct implementation of strategies from Bloom's (1968) theory of mastery learning.

The key factors of Bloom's (1968) work of increasing the instructional time period, providing personalized instruction for undemonstrated learning objectives, and giving students an opportunity to raise the grade from a retake, or re-demonstration of knowledge are what differentiate mastery-learning approaches from other grading practices that do not allow additional opportunities to demonstrate increased learning (see Figure 2). These factors also confirm the selection of learning-for-mastery as the conceptual framework for this research study, which evaluated the correlational analysis of the accuracies of traditional total-points versus assessment-based grading policies. For the purpose of this study, it is assumed that mastery-learning was not purposefully incorporated into a total-points grading practice. While each classroom teacher generally has some flexibility within any mandated grading practice, total-points grading is assumed to follow a pattern of instruction then assessment without the specific steps of reteaching and reassessment that differentiate mastery-learning practices. In this study, total-points grading will not use mastery-learning strategies, but assessment-based grading does incorporate the key components of Bloom's (1968, 1974, 1976, 1982, 1984) theory of mastery learning because it included an individual reteaching session before the additional summative assessment opportunity (see Figure 2).

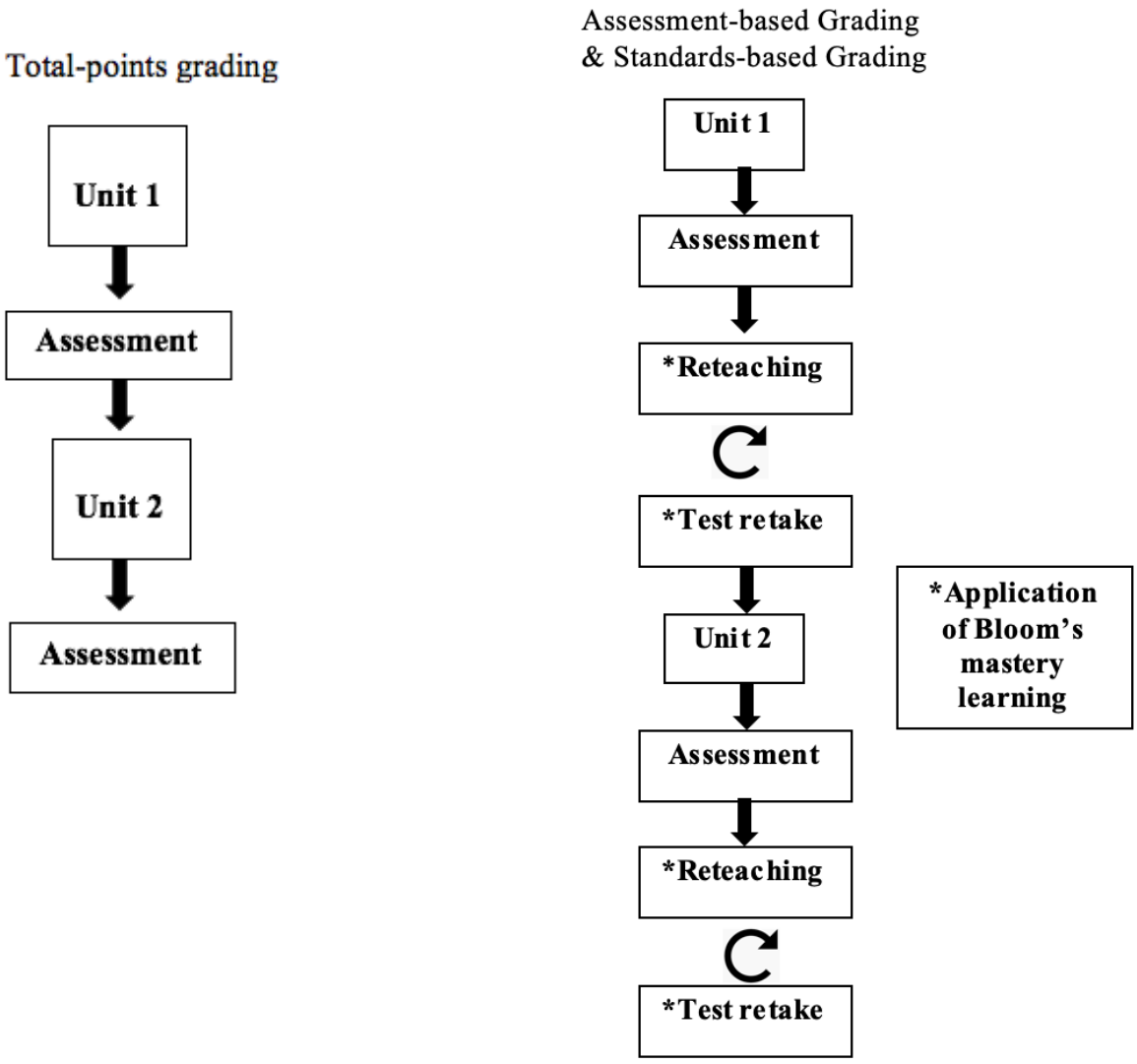

Figure 2. The role of mastery-learning strategies in assessment-based grading practices (Winter, 2019).

An outcome-based grading system heavily weights summative assessment grades and provides numerous opportunities to demonstrate improved mastery. As an outcome-based grading system, the ultimate goal is to maximize student learning through instructional and assessment practices. It is important to recognize standards-based grading as a bridge for the process of mastery learning as it is applied in assessment-based grading. The practice of

standards-based grading shifts the focus from an averaged numerical representation of student performance for comparison, to a summary of student mastery of the content learning objectives. A report card issued using a standards-based grading practice, generally used rubrics with values to summarize student performance on a learning objective. On the rubric, students would be assigned a proficiency score of one through four; one indicating developing skills through three proficient and four indicating mastery. Other report card formats using a standard-based approach, or in feedback to the students, the learning objectives were categorized and listed as either still needing to be mastered, or already mastered (Marzano & Heflebower, 2011). In a standards-based grading practice, students are continually assessed to check for mastery. Any improved student performance, demonstrating greater mastery, supersedes any previous inferior performances and proficiency score. The standards-based practices of continuous assessment, followed by instructional intervention and reassessment, are also used in assessment-based grading.

The role of standards-based grading can be summarized as an assessment for learning instead of an assessment of learning (Wormeli, 2009). Like assessment-based grading, standards-based grading practices also aimed to eliminate the inaccurate inclusion and influence of student behaviors on the academic grade (Wormeli, 2009). Essentially, assessment-based grading was an academic grade assigned in the representation of standards-based learning indicators, because the assessment-based score as so heavily weighted to emphasize student summative performance on the described learning objectives (O'Connor, 2012). This extended and continuous feedback loop is the heart of standards-based and assessment-based grading practices. An important factor in the standards-based grading framework is the extended opportunities for students to demonstrate increased content knowledge and mastery. Marzano

and Heflebower (2011) suggested that assessment retakes or re-grade opportunities occur continually throughout different grading periods. This means that students would have the opportunity to improve on assessment scores. Likewise, a key component of the assessment-based grading practices is the opportunity for students to retake assessments to improve scores and promote increased mastery (O'Connor, 2012). A shared benefit of both standards-based and assessment-based practices is the students' potential instructional gains from a mastery-oriented instructional and grading approach. Consequently, the theory of mastery learning served as the appropriate conceptual framework to support the usage of both standards-based and assessment-based grading practices.

It is important to note that Bloom (1982), also recognized the compounding effect of success and failure as the material scaffolds in sequential learning units. A traditional total-points grading practice, when viewed through the lens of a mastery-learning approach, was susceptible to being less effective overall because of the limited learning and testing window offered to students. Offering summative test retake opportunities and extending the instructional time, both components of assessment-based grading, were more supportive of mastery learning. The weighted category grading emphasis used in assessment-based grading clearly stresses a prominence on student-demonstrated knowledge of the learning objectives, which is the purpose of mastery learning. It is beneficial to facilitate as many students meeting the mastery level as possible.

The diverse skillsets required to be successful in the contemporary work force necessitates a variety of employee strengths and experiences. It is no longer beneficial to exclude groups of students, assuming that only top-tier students will be contributing to the workforce in a meaningful way (Block et al., 1968, pp. 47–48). For our society to stay

competitive on a global level, educators cannot afford to ignore and exclude groups of students. Mastery-learning increases the size of the successful student population who will be better prepared to continue their education or contribute to the workforce as more skillful employees (Bloom, 1968).

This study sought to investigate if and how an assessment-based grading approach is more accurate of cumulative student content knowledge than a total-points grading system. Standards-based grading practices use a mastery-driven instructional approach and proficiency rubrics to evaluate student learning. The literature review presents some conflicting results in the accuracies of grades from standards-based practices. Discrepancies in the current literature for the support or opposition on standards-based grading facilitate an opportunity for this research to provide more evidence in the outcome-based grading discussion. Outcome-based grading practices include both standards-based and assessment-based grading because of the emphasis on cumulative, summative learning. In this study, the grading practice that more closely matched student standardized test score values as a benchmark indicator of content mastery of learning objectives can be deemed a more accurate representation of comprehensive student knowledge. This research may provide insight into best practices in classroom grading methods.

Review of Research Literature and Methodological Literature

The methodological support for the practice of assessment-based grading stemmed from three major bodies of research, including the literature regarding the effectiveness of standards-based grading, assessment-based grading, and total-points grading. Furthermore, there is a body of research that supports learning for mastery techniques as a best practice. Assessment-based grading was a derivative, essentially the numerical and letter-grade representation of the

descriptive categories of standards-based grading. The utility and benefits of a standards-based grading approach was well documented through research (Beatty, 2013; Brophy, 2013; Knaack et al., 2012; Pollio & Hochbein, 2015; Rosales, 2013; Timmons, 2017). There was specific research that directly supported the usage of assessment-based grading practices in the classroom (Pekel, 2015; Saefurrohman & Balinas, 2016). Other researchers supported the usage of a traditional total-points grading strategy (Akins, 2016; Comes, 2015; Denson, 2013; Fink, 2015; Lee, 2013). The two homologous grading practices, assessment-based grading and standards-based grading, are based on instructional practices geared towards mastery learning. As an older theory, there was ample research in support of the success of Bloom's (1968) learning-for-mastery approach, dating back almost five decades, starting in the 1970's (Mitee & Obaitan, 2015; Smeding, Darnon, Souchal, Toczek-Capelle & Butera, 2013). These four avenues provide a synthesis of current research to provide the appropriate context for best examining the accuracy of assessment-based grading in comparison to a total-points grading system.

Research addressing the accuracy of standards-based grading. Standards-based grading is similar to assessment-based grading because of the heavy emphasis on student demonstration of the learning objectives. A student score is assigned, based on individual competencies of content learning objectives. There is a pool of research that supported grade calculation based primarily on student mastery of the learning objectives. Pollio and Hochbein (2015) conducted a research study that identified the relationship between standards-based grading and standardized test scores for high school mathematics students. Study participants included Algebra 2 students from 11 high schools in Kentucky, with a cohort of 2,419 students. Students were graded in the classroom using a standards-based grading approach and all participating students took the Kentucky Core Content Test (KCCT) in mathematics at the end of

the year (Pollio & Hochbein, 2015). The performance of students on the math KCCT was categorized into mastery categories as novice, apprentice, proficient, or distinguished. Classroom letter grades using a standards-based grading approach classified students into the following mastery categories: A (exceeds standards), B (meets standards), C (marginally meets standards), D (below standards), and U (unsatisfactory performance) for three mathematics standards over a six-week period (Pollio & Hochbein, 2015).

Researchers used a quasi-experimental non-equivalent control group design to analyze the relationship between standards-based grades and standardized math test scores for the 2011 school year. The first control group included students that completed Algebra 2 that took the KCCT math test the previous school year, but with a total-points classroom grading method. The second control group included the same cohort of students that received the standards-based grading in 2011 in Algebra 2 and took the 2011 KCCT math test, but received total-points grading in their science class (Pollio & Hochbein, 2015). This non-equivalent control group reduced the greatest threats to validity based on students' efforts. Pollio and Hochbein (2015) used analysis to find the correlation coefficient for each group between the standards-based grades of students and KCCT test scores. The group with the standards-based math grades was found to have the highest correlation to KCCT math test scores.

Pollio and Hochbein (2015) used a descriptive statistical analysis of data from 2010 math students that experienced traditional total-points grading, and found a weak relationship between classroom-assigned grades and KCCT proficiency; only 26% of the 440 students who earned As and Bs in the total-points grading model earned a proficient or distinguished rating on the 2010 KCCT math test, although these students represented 40% of the student population. About 55% of the 568 highest-achieving students, which consisted of 45% of the student population, who

earned an A or B in the standards-based grading method also received a score of proficient or distinguished on the 2011 KCTT math test (Pollio & Hochbein, 2015). The results from this study indicated that when teachers switched from traditional total-points grading to standards-based grading, the number of students who earned A and B classroom grades increased, as did the rate of passing the KCTT state math assessment (Pollio & Hochbein, 2015). The researchers' efforts to include two control groups and a large population size increased the validity of the study. This pivotal point for the district, where all 11 high schools transitioned from total-points grading to standards-based grading in math classes, provided a meaningful opportunity for data analysis of the relationship between classroom grading practices and test scores. The study by Pollio and Hochbein (2015) provided research-driven evidence in support of standards-based grading practices.

Hardegree (2012) conducted a quantitative research study with a non-experimental causal comparative design that focused on the variance in standards-based grading report cards compared with standardized test scores. In a qualitative study, Hardegree (2012) provided an analytic comparison of the standards-based grade cards, or SBRC, of 550 fifth-grade students in Georgia as opposed to individual standardized test scores from a criterion referenced competency tests in math and reading. Both evaluation measures categorized students into categories if they did not meet, met, or exceeded the standards. The results of the study showed that the students scored higher on the standardized math test than on the SBRC; the SBRC underestimated the test scores for most students (Hardegree, 2012). Data analysis of the student reading scores using the P-value of one-way ANOVA model was smaller than 0.0001, which indicated that there was some alignment between the SBRC and test scores for all students. However, discrepancies were found in student reading scores and test values, providing evidence of "significant differences"

(Hardegree, 2012, p. 97). Notably, Hardegree (2012) also found that students on free/reduced lunches scored lower on the standardized math test, even if they had comparable SBRC scores. The study acknowledged some alignment between the standards-based grading practice and student test scores, but calculated statistical variance that warrants further research but supports the overall presumption that standards-based grading can provide accurate information on student learning and act as a reliable indicator of student performance on standardized tests (Hardegree, 2012).

Beatty (2013) presented data that standards-based grading is also successful for students at the postsecondary level. Beatty (2013) explained how standards-based grading was a more accurate representation of student learning, promoted deeper understanding and learning even after initial failures. The research represents a range in student ages, demonstrating the overall effectiveness of standards-based grading and its applicability in all levels of education. Timmons (2017) researched students' perspectives on the usage of the standards-based grading method by surveying English II Honors students using a Likert scale and further interviewing five from the student population. Timmons (2017) presented the 10th grade English Honors II students with a report card including both a single numerical average score and standards-based grading report summarizing a four-and-a-half week grading period in the fall semester 2016. The results of the research study stated that students, as a whole, preferred the standards-based reporting to the numerical value and viewed the standards-based evaluation tool as accurate and fair (Timmons, 2017).

Rosales (2013) conducted a quantitative research study with a causal-comparative design that evaluated the effect of standards-based grading as a treatment on summative student test scores. Two high school algebra teachers each taught two sections of Algebra 2. All students

were given a multiple-choice pretest. One section from each teacher, or two of total four sections of students, was taught using a standards-based grading approach for the 12-week session and the other two sections, one from each teacher, was taught using a traditionally graded total-points assessment approach without the opportunity for summative retakes. The students in the standards-based grading sections were provided with weekly opportunities to retake assessments until each individual reached mastery standards on the learning objectives and continually feedback on mastery levels. The two instructors planned lessons together weekly to ensure that each section was provided with the same practices and opportunities. At the end of the 12-week session, all students were given the same posttest. The descriptive statistics compared the posttest mean scores for students in the treatment and control group and showed similar score means for students taught in a standards-based method versus the traditional total-points grading. These results indicated the selection of a standards-based grading practice had no advantage or increased students' mastery as demonstrated on a cumulative posttest over a traditional total-points grading approach (Rosales, 2013). The fact in this research study the grading method was not found to be an influential factor on student test score increases the need for further research (Rosales, 2013).

Not all research found a strong positive correlation between teacher-assigned, standards-based grades, and student test scores. Welsh, D'Agostino, and Kaniskan (2013) conducted a study to measure the correlation between standards-based grades and student test scores. All third-grade and fifth-grade students from 11 schools were included for a study population of approximately 750 students each year, over a 2-year period for a total sample size of 3,026 students (Welsh et al., 2013). At the time of data collection, the participating district had implemented standards-based grading for three years. Standardized test scores, disaggregated

into content categories, were matched to corresponding second-semester standards-based report cards progress to evaluate the convergence. A coefficient kappa, Tau-b correlation, and mean difference tests were used for statistical analysis and disaggregated by Grade 3 and Grade 5 in Year 1 and Year 2 for the subjects of mathematics, reading, and writing (Welsh et al., 2013). Kappa values ranged from $k = .188$ to $k = .321$ and documented only slight level of association. The Tau-b correlations were somewhat stronger but still only indicated a moderate relationship at best. The degree of association between standards-based grades and test scores was significantly different from zero for all analysis ($p < 0.05$). Based on these results, researchers found only a moderate to weak correlation between standards-based grades and test scores (Welsh et al., 2013).

Brophy (2013) focused on the impact of student motivation, specifically resulting from retake opportunities, to increase mastery. Brophy (2013) found that student motivation was related to grades and the amount of content mastered. Awarding academic grades, specifically with an opportunity to improve them, supported a behaviorist's approach to teaching and provided feedback for students, which increased student motivation to work toward the desired outcome of mastery. These factors of the goal theory are conducive to the tenets of mastery learning. Brophy (2013) also found that a teacher's adverse view on a student or his or her behavior, attendance, and student personality could, consciously or subconsciously, negatively influence the assigned academic grade. This finding confirmed the summative grade weighting in assessment-based grading practices and worked to mitigate the potential of a teacher's negative view on student behavior to negatively influence a student grade.

Similar to Brophy (2013), Knaack et al. (2012) agreed a grade that represented content mastery was the best representation of the student. Teachers, students, and parents were

interviewed to gain insight into stakeholders' views on traditional grading practices. The data-collection process included 158 students, 95 parents, and 14 teachers. Students and parents shared their concerns that the previously assigned academic grade was influenced by non-summative scores and did not convey student mastery of the material because of the inclusion of so many other scores (Knaack et al., 2012). Teachers then changed to a more standards-based grading approach and used summative grading as a feedback mechanism to identify areas of potential growth for students and provided students with opportunities for improvement. With these grading changes, researchers found that 84% of students reported that they agreed with the fairness of the grades assigned to them and 92% of the parents interviewed understood the student grades (Knaack et al., 2012). This study documented the student and parent preference for a standards-based grading system that represented student mastery of the learning objectives.

This body of research provided evidence-driven support for the usage of standards-based grading (Beatty, 2013; Hardegree, 2012; Pollio & Hochbein, 2015) as an accurate representation of student learning in comparison to test scores, and provided data that demonstrated student and teacher preference for standards-based approaches. Brophy (2013) provided a link between continued opportunities to demonstrate mastery and increased student motivation. However, Rosales (2013) and Welsh et al. (2013) challenged the strength of the correlation between standards-based grades and student test scores. These results can be viewed as a supplement to the research regarding the accuracy of assessment-based grading.

Research addressing the accuracy of assessment-based grading. Assessment-based grading is a derivative of standards-based grading because the primary composition of the class academic grade is based on student performances of the targeted learning objectives. Saefurrohman and Balinas (2016) used a mixed-methods approach to research to evaluate the

effectiveness of assessment as a feedback mechanism during instruction. Assessment as a continued tool for learning is a key component of the assessment-based grading practice. The study sought to explore a trending shift from assessment of learning to assessment for learning (Saefurrohman & Balinas, 2016). The study evaluated the numerical data from 48 teacher questionnaires and 12 teacher interviews that were centered on classroom assessment practices.

This transition assessment of learning to assessment for learning changed the role of summative assessments, from providing an end of unit evaluation to offering a working diagnostic to provide feedback to learners for continued growth. Researchers found that teachers who used classroom assessment for learning did so to group students for further, more individualized instruction, to diagnose strengths and weaknesses in teacher instruction, and to provide feedback for students on specific areas for improvement. Assessment became a step in the feedback cycle, not an ending point after instruction (Saefurrohman & Balinas, 2016). The retake component of the assessment-based grading approach, or opportunity to increase learning and mastery then demonstrate that newly gained knowledge, was conducive to a mastery learning approach because it elongated the instructional period, provided individualized feedback after an evaluation with specific reteaching and an opportunity for students to further demonstrate their content knowledge through a summative assessment. These research studies and empirical reviews provide evidence-based support for an assessment-based grading approach that stems from the corresponding standards-based grading.

Pekel (2013) designed a study that compared eighth-grade students' grades calculated with a traditional total-points system to state standardized math scores and assessment-based derived student grades to state standardized test scores. In 2007–2008, the correlation coefficient between traditional total-points student grades and state standardized test scores was nearly

negligible at 0.00 to 0.20 (Pekel, 2013). In the district's grading reform, teachers changed to assessment-based grading practices that removed the influence of nonacademic or noncognitive factors and followed the stipulations that student grades consisted of weighted categories; homework had a 0% weight in the overall grade. Additionally, students also had the opportunity for test retakes on summative assessment scores. Following the intervention of assessment-based grading practices, the correlation coefficient between the assessment-based scores and state standardized test scores rose to the range of 0.40 to 0.60 (Pekel, 2013). The data analysis, showing a stronger correlation between assessment-based grades over total-points grades, provided evidence that supported assessment-based grading as an effective practice in terms of accuracy when measured against standardized test scores.

O'Connor and Wormeli (2011) cited previous grade reports in their literature review that provided ample documentation to show how averaging grades falsified grade reports and failed to show content mastery disaggregated by standard. The inclusion of zeros in a grade book creates a similar skew that leads to a misrepresentation of student learning (O'Connor & Wormeli, 2011). To increase the accuracy and effectiveness in grading, teachers need to separate nonacademic elements from academic elements on the report card. Factors that should not influence a grade include: nonacademic acts, like an organized notebook that doesn't align with a learning objective or extra credit, group work, problem with averaging test retakes, the skew of a zero in the grade-book (O'Connor & Wormeli, 2011). Researchers suggest that instead, teachers assigned a grade based on performance standards to increase the consistency and accuracy of the grades assigned (O'Connor & Wormeli, 2011). Accurate grades provide valuable feedback and promote continued learning. Research has shown that an assessment-based grading practice is more aligned with a mastery-learning approach to instruction than the

traditional total-points grading system (Pekel, 2013). This research sought to assess if an assessment-based grading practice can use the research-proven positive outcomes of a mastery-learning approach and assign a numerical value to a derivative of a standards-based grading practice. There is substantial research-driven support for a mastery-oriented approach, as summarized by the subsequent research findings later in Chapter 2.

Research addressed the accuracy of traditional total-points grading. Denson (2013) conducted a correlational, non-experimental research design to examine the relationship between teacher-assigned grades and students' test scores. The researcher used historical data in the form of 10th-grade math students' classroom grades, as determined by a total-points method, and math section scores from state's standardized test, Florida Comprehensive Assessment Test, or FCAT, for the 2009–2010 and 2010–2011 school years in the spring semesters (Denson, 2013). The researcher disregarded ethnic subgroups groups with fewer than 10 members and excluded students that were not properly tested, resulting in a convenience sample of 273 students that met the inclusion criteria. Denson (2013) used the Pearson correlation analysis identified a relationship of statistical significance between students' total-points classroom grades and standardized test scores on the math FCAT ($r = 0.41$, $p < 0.001$). Descriptive statistics using a multiple regression model controlling for student characteristics like gender and ethnicity also confirmed a significant relationship between student total-points, teacher-assigned math grades, and FCAT standardized math scores. Denson (2013) confirmed the effectiveness of using a total-points grading strategy, as it accurately reflected students' classroom proficiencies comparable to students' performances on the FCAT. The convenience sample with subgroup breakdown and descriptive statistics increased the credibility of the study.

Fink (2015) designed an action research study to compare a standards-based grading practice to a traditional, total-points approach in respect to student achievement. Two high-school teachers volunteered to participate with their combined 63 students. Each teacher had a class that used traditional total-points grading, and a different class to which they applied a standards-based grading approach with retake opportunities. On average, the students in the traditionally graded sections had a higher class-average than those classes with standards-based grading at 80% and 76% respectively, (Fink, 2015). In terms of academic achievement as represented by a semester grade, students in the traditionally graded total-points sections had, on average, a greater academic achievement. Student interviews surveys, however, reported an increased student preference for assessment retake opportunities in the standards-based approach (Fink, 2015).

Comes (2015) conducted an action research study to compare traditional total-points grading as opposed to a criterion referenced system, or standards-based grading. Eight sections were in the study design of sixth-grade science classes consisting of 179 students total; a stratified sample of 60 students was used for the study population. Four of the sixth-grade science class sections used criterion referenced grading, or CRG, and the other four sections were graded using a traditional total-points system. There were two participating teachers, and each teacher had two classes with each grading practice (Comes, 2015). Students were administered a pretest and posttest. Student records from each grading system were compared to posttest scores to determine which grading system was a more accurate predictor of students' posttest performances (Comes, 2015). Through a Pearson correlation analysis of the student data, Comes (2015) demonstrated a weak correlation of 0.314 that did not warrant statistical significance between the classroom grades derived from the CRG method and the students'

posttest scores. However, the researcher found a strong positive correlation with a median correlation of 0.726 between the traditional total-points classroom grades and the posttest scores of students. Essentially, the data analysis supported a traditional total-points grading system as a more accurate predictor of students' scores on posttest performances (Comes, 2015).

Lee (2013) also evaluated the accuracy of traditional classroom grades of students in correlation to standardized test scores. A correlational mixed-methods approach sought to identify the relationship between teacher-assigned grades and state standardized test scores, as measured by the Arizona Instrument to Measure Standards (AIMS). The study population was derived as a representative stratified sample based on the socioeconomic status of students with 65 students from each demographic rating with a total student population of 230 elementary students (Lee, 2013). The researcher applied a Spearman rho test to the student data and identified a strong positive correlation, with a correlation coefficient of .82, between teacher-assigned classroom grades and standardized test scores (Lee, 2013). This provides evidence of the accuracy of traditional total-points classroom grading practices.

The longevity of use of traditional total-points grading practices makes it a comfortable choice for many educators (O'Connor, 2009). Research by Akins (2016) demonstrated the preference for a traditional total-points grading system by high school administrators. Akins (2016) conducted a descriptive, non-experimental quantitative survey of 247 lead high school principals from across the state of Missouri to evaluate themes in perception about the traditional total-points grading practice. Participants were given an attitudinal survey to assess the perceptions of the traditional grading practice from secondary principals. Descriptive statistics were used to identify themes in the responses. Of those provided with the survey across the state of Missouri, 33% agree with traditional grading practice of determining a term grade through

averaging assignments (Akins, 2016). The principals were divided in their reported support for grading practices included in the traditional total-points approach. Around 35% of principals reported that late-work should be penalized through a point deduction and 40% instruct teachers to enter scores of zero for missing work (Akins, 2016). The results from this study showed the preference from high school principals for the utilization of a traditional total-points grading practice over other available grading procedures.

Although the research from Comes (2015), Fink (2015), Lee (2013), and Denson (2013) does support the use of traditional grading practices, there is research that challenges the accuracy of teacher-assigned grades using traditional grading practices. Merki and Holmeier (2015) conducted a 5-year longitudinal study to evaluate the relationship between standardized achievement test scores and semester grades of students. Students were administered the Trends in International Mathematics and Science Study exam to serve as the standardized test benchmark. The study population consisted of 1,288 secondary students from 13 schools, with a response rate of 55% to 71% for advanced placement math courses but only 16% to 24% response rate for basic math courses. A correlational analysis was conducted for each year to determine the relationships between standardized math test scores, classroom semester grades derived from a total-points system, and end of the year content exit exams.

Merki and Holmeier (2015) found a moderate positive correlation between standardized math test scores and semester classroom math grades of students derived from a total-points system (r = .42). Analysis demonstrated the strongest correlation between math grades and standardized test scores, with weaker relationships in biology (r = 0.36; n = 776) and English (r = 0.30; n = 1038) respectively (Merki & Holmeier, 2015). Further analysis documented the strongest correlation between classroom math grades and standardized math test scores for

advanced students. This finding confirmed concerns voiced by proponents of standards-based grading that traditional total-points grading practices may be affected by the inclusion of nonacademic factors, especially in remedial courses. Merki and Holmeier (2015) used statistical analysis to demonstrate a stronger correlation between standardized test scores and content exit exams than between standardized test scores and classroom grades. The inclusion of content exit exams for analysis, however, provided insight into the academic validity of the classroom semester-grades in relation to a single summative assessment score. Based on the 2007 scores, there was a relatively low average correlation of $r = .34$ between standardized math test scores and classroom exit-exam scores than found with classroom grades at the 13 participating schools (Merki & Holmeier, 2015). Furthermore, researchers noted a large variance in correlations between different schools; it was easier to earn better grades at some schools than others with the same test performance values (Merki & Holmeier, 2015). The weak to moderate correlation values derived from the statistical analysis provided in this research acknowledged the potential inaccuracies between traditional classroom-assigned total-points math grades and standardized math test scores (Merki & Holmeier, 2015).

In the study by Merki and Holtmeier (2015), participating teachers used a total-points grading approach, but each had different point-assignments to different categories and assignments resulting in inconsistent classroom grade determinations. This is a susceptibility for all total-points grading practices, as there are no set recommendations for the proportion of point-assignments and each teacher applies his or her own approach to point-assignments which consequently weights different assignments within the total-points system. In a total-points grading system, the points available for participation and homework is usually enough for low-achieving students to earn a passing grade (Akins, 2016). In a total-points system, students that

are responsible and have a good work ethic may earn a decent or at least passing grade in a class without demonstrating high levels of learning of the learning objectives and standards (Akins, 2016).

Research confirmed students learned more through mastery learning methods. Mitee and Obaitan (2015) specifically studied the effect of mastery-learning practices on students' cognitive learning outcomes. Researchers used a quantitative approach with an experimental design with a test treatment group and control group. There was a sufficient sample size (n = 401); a control group was included and the same pretests and posttests administered to each group (Mitee & Obaitan, 2015). Researchers divided 401 secondary chemistry students into a treatment group with a learning-mastery approach and control group, using traditional instruction without the assessment and then individualized remediation and reassessment components. The mastery group outscored the control group on a posttest. Of the mastery students 69% scored an 80% or higher on the test and only 17% of the students in the control group hit the same marks (Mitee & Obaitan, 2015).

The summative test scores showed that students demonstrated higher achievement through higher overall scores on a final chemistry exam when mastery-learning instructional techniques were used when compared to traditional methods in the control group. Given more time and individualized instruction, the students learned a larger percentage of the learning objective and were more academically ready for new scaffolding information (Mitee & Obaitan, 2015). This research study confirmed the effectiveness of a mastery-learning approach in summative student achievement in a secondary setting. The study did not mention if one group covered more content, or a greater number of learning objectives, than another.

Smeding et al. (2013) conducted two studies that demonstrated how a mastery-oriented approach can be used to help decrease the achievement gap between low-socioeconomic (SES) students and high-socioeconomic (SES) students at the postsecondary level. In the research design, 246 first-year psychology students were divided into an experimental and control groups. Mastery-approach techniques with continuous assessment were applied to the experimental group. Students were continually assessed and provided with specific learning objectives. While high-SES students still outperformed low-SES students overall, the low-SES students participating in the mastery-based section outscored their counterparts in the traditional instruction design with performance-based examinations comparatively on the final exam (Smeding et al., 2013). An analysis of final test scores showed that a mastery-oriented approach supports low-SES students' achievement and is important for decreasing the achievement gap between SES groups at the postsecondary level.

A second phase to the research was designed to measure the effect of student mastery mindset on summative academic performance. Researchers set up an experimental and control group to try to isolate the effect of the treatment, or mastery-oriented approach (Smeding et al., 2013). The researchers ran a mixed analysis of covariance with SES as the between-participants variable. Researchers designed a subsequent study with 233 first-year psychology students to evaluate how students' views on mastery goals related to grade outcomes. An analysis of grades found that while again high-SES students outperformed their low-SES peers, the low-SES students who exhibited a high mastery goal endorsement significantly outperformed students that denoted a low mastery goal endorsement level on a final psychology exam. The results demonstrated a high statistical significance between SES and mastery goal interactions (Smeding et al., 2013).

The achievement gap is reduced between high-SES and low-SES groups when a mastery approach is used for course design and instruction or also reduced for students that identified as endorsing a mastery-oriented goal-setting approach (Smeding et al., 2013). The research design was directly aligned to the research questions and provided research-based evidence to support a mastery-oriented instructional approach and fostering of a high mastery goal endorsement for students. This research study provides clear evidence to support an instructional approach that supports the ideals of mastery learning at the postsecondary level (Smeding et al., 2013). Therefore, this research will seek to discover if there are similar benefits for low-SES secondary students in a mastery-oriented approach through assessment-based grading practices. Smeding et al. (2013) is related to this research because it evaluated the effect of a mastery-oriented instructional approach on summative student learning. Likewise, this research study used a similar approach, but with a summative class grade and standardized test score as the comparative measure instead of a final test score. Smeding et al. (2013) is related to this research study because it evaluated the effect of a mastery-goal or growth mindset in students on academic performance. This message is not directly taught in an assessment-based approach, but is modeled through the test-retake allowance. Smeding et al. (2013) is related to this research because it isolates student SES as a factor that should be analyzed in this research data analysis, and the population for this study has a majority of low SES students.

The problem statement addresses a gap in research. Saefurrohman and Balinas (2016) and Pekel (2013) conducted mixed-methods and correlational research to evaluate the effectiveness of assessment-based grading. O'Connor and Wormeli (2011) conducted an empirical review that synthesized the results of a few previous studies; overall there was limited research available specifically addressing the accuracy of assessment-based grading practices.

Many school districts in the district's state have implemented an assessment-based grading approach in the last five years in alignment with O'Connor's (2012) description of assessment-based grading. The Midwestern school district that provided the setting for this research adopted O'Connor's (2012) assessment-based grading policy in 2013, as mandated by the building administrators (XXXXX & XXXXXXXX, 2017). The very small body of research available on assessment-based grading by Pekel (2013), O'Connor and Wormeli (2011), and Saefurrohman and Balinas (2016), provided an opportunity for meaningful research to evaluate the relationship between assessment-based student grades in relation to standardized test scores in comparison to comparable grades derived from the traditional total-points grading method.

Review of Methodological Issues

Despite sharing similar topics, researchers selected different research methodologies as an approach to conduct research on grading practices. There are strengths and weaknesses to each methodological design. The instructional techniques and grading practices summarized in the literature review showed that most researchers selected a quantitative approach. More specifically, an experimental design as the quantitative approach was a frequent research methodology for the previous research selections. Other researchers used a correlational design to identify and measure the relationship between two variables.

Pollio and Hochbein (2015), Pekel (2015), Denson (2013) and Lee (2013) all used a quantitative approach with a correlational design. The descriptive research method was used to analyze the relationships between student grades from different grading practices compared to test scores of students. Rosales (2013) designed a quantitative causal-comparative approach to grading analysis. The methodology was correlational because the researchers did not manipulate any variables. This was the best study design choice for the parameters of the data range used.

Each study had different populations sizes and resulted in varying strengths in correlational relationships, but most closely mirrored the methodology in this quantitative correlational research study.

Smeding et al. (2013) conducted two quantitative studies with an experimental design. The hypothesis from researchers for the second study was based on the results for the first portion of the study in a randomized field experiment. The study also included a component, assessing student motivation that used a 7-point Likert scale survey as a measurement tool and ANOVA statistical evaluation for variance (Smeding et al., 2013). Saefurrohman and Balinas (2016) also used a mixed-methods approach with teacher questionnaires and analysis of student data to specifically evaluate the effectiveness of assessment-based grading. The quantitative aspect used a survey with a statistical analysis. The researchers also used a questionnaire, interview, and observation of different classroom teachers at different high school buildings. The quantitative data were analyzed using descriptive and inferential statistical models, and the qualitative data was analyzed using descriptive analysis and applied as a theme to the research summary. A concern with the interviews and researcher observations is the potential for researcher bias, as with all research, and is acknowledged by the researcher. Lee (2013) used a mixed-methods approach including correlation analysis and participant surveys. The methodology in a mixed-method study used a multi-technique approach that confirmed the soundness of the results.

Hardegree (2012) conducted a quantitative research study with a non-experimental causal-comparative study. The standards-based scores for more than 500 students were converted into a numerical scale and compared to individual standardized test results. The selection of already established data minimized the risks of researcher bias because the

researcher was not involved with the collection of the data, only the analysis. Rosales (2013) also used a causal-comparative quantitative design. This research design choice was more appropriate than a correlational research approach because the researcher did not have the liberty to randomly assign students to the treatment and control groups. Instead, the school counselors assigned students during the scheduled enrollment process. The convenience sample for the study consisted of approximately 10% of the student population at the school, 107 students participated in the study. The researcher conducted a sample t-test on the teacher written pretest, which was proctored under consistent conditions, to ensure that all sections were of comparable ability at the start of the study (Rosales, 213). This was important in establishing a standards-based grading strategy as the variable and isolating any resulting effect as attributable to the treatment and not caused by advantages in the student population section assignments. The same students were given the same posttest and a two-way ANOVA analysis was applied to look for any significance differences in the mean final test scores of students in different sections (Rosales, 2013). This was a suitable choice because the research questions centered on investigating the interaction between two independent variables on one dependent variable. The researcher was one of the instructors whom taught two of the four sections in the study, which raised some concerns about researcher bias. However, it is evident from the meticulous methodology and study design that all efforts were made to mitigate this influence and the researcher addressed this concern.

Knaack et al. (2012) executed an action-research project, which is a common choice for a school setting because the practitioner conducted the research. The research project included an initial survey of students and parents, a change in the grading practice to a standards-based approach, and post-term survey to assess the effectiveness of the change. The data were

collected from students and parents of three different teachers with a total student population of 253 students and parents. This research design is susceptible to the same weaknesses as the previous study due to the concerns with self-reporting surveys but does benefit from the practicality of action-research in a classroom setting. Fink (2015) and Comes (2015) also conducted action research studies to evaluate grading practices. These studies that used a traditional total-points grading system did not regulate the grading practices to ensure that all teachers used comparable grading strategies and weights, allowing for variance in teacher grade assignments that cannot be accounted for in the statistical analysis. Nevertheless, there are strong advantages of an action research methodology, especially in an educational setting because the topics are often meaning to teachers, directly applicable to instructional effectiveness, and works to facilitate continual school improvement.

Mitee and Obaitan (2015) used a quantitative approach with a quasi-experimental design with a test treatment group and control group. Researchers administered a 25-question pretest and posttest; the data analysis used descriptive statistics and a sample t-test. This was an experimental approach because the researchers attempted to regulate all controls except for the variable tested. The study sample population was pulled from four different secondary schools, which increases the confidence in the results. The large random study population from multiple sources increased the applicability of the results across larger populations. A concern with this quasi-experimental design is the ability to control all factors, except the instructional strategies. The instructional techniques were self-reported by the teachers and may or may not have aligned with true mastery techniques. Smeding et al. (2013) also used an experimental design with control and experimental groups. An experimental design methodology can provide strong

evidence for the effectiveness of a treatment, but it can be difficult to secure a meaningful population size and comparable treatment and control groups.

Two of the researchers used designs with a qualitative approach. Beatty (2013) used a qualitative case-study design to determine if standards-based grading was a more accurate representation of student learning than non-mastery methods. The researcher made observations and reported data from class assessments and student self-reported views via a questionnaire. There was no plan for student assessment and the results are based on a self-assessment from the instructor in his own course. Timmons (2017) also used a qualitative approach but with an action-research format. The quantitative data were collected through a Likert scale and student interviews reporting perspectives from students on grading practices. As with the previous research designs using surveys, Likert scales or other self-reporting mechanisms all fall prey to the same concerns about accurate and honest reporting by the study participants. Similarly, Brophy (2013) used surveys to determine student motivation in relation to mastery approaches and teachers' perceptions of students that influence grading. Akins (2016) took a similar approach to solicit administrator feedback on grading practices.

The current study used a quantitative approach to identify which grading practice is a more accurate representation of learning for eighth-grade math students, as compared to a benchmark of national standardized math test scores and will measure the strength of the correlation through an analysis of the data. This research used a quantitative approach because it will evaluate and measure the strength of a relationship between variables using a numerical statistical analysis through a correlational study. This design approach presents the best analysis of the data available, including individual student scores. A concern with all quantitative studies, which this research was likewise susceptible, is the opportunity for researcher bias. Researcher

bias is an acknowledged concern for this research study. To mitigate these concerns, the researcher used student scores from another teacher. Additionally, grading was assumed to be as objective as possible. The assignments and tests administered to students were included in the school-provided curriculum. The same scores, accommodations, and modifications were used for each grading practice so that the only variable was the application of the grading policy weighted categories. These steps created protections against personal bias. Therefore, a correlational quantitative design is the most appropriate methodology for this research.

Synthesis of Research Findings

There are common themes found throughout the literature review. Previous research clearly supports a mastery approach for increased student learning and both standards-based (Rosales, 2013) and assessment-based grading practices (Saefurrohman & Balinas, 2016) have been shown to promote a mastery-oriented approach. Brookhart (2011) maintained that the primary issue with grading is not a matter of grading scale, frequency of assessment or the weight and combination of scores; the essential grading question is what do we want our grades to convey? Who is the intended audience for the message? Standards-based grading is based on the principle that grades are about what students learn, not what they earn; a grade should convey achievement of corresponding standards (Brookhart, 2011). Evidence-based data supported learning-mastery instructional approaches in the classroom, documenting the effectiveness of standards-based grading and the promise of assessment-based grading practices (Mitee & Obaitan, 2015; Smeding et al., 2013). Parents and teachers alike self-reported an increased preference for standards-based reporting form over the traditional form by a wide margin (Brophy, 2013). However, on a survey, Akins (2016) found that administrators are most familiar

and comfortable with traditional grading methods over standards-based or assessment-based practices.

The selection of standards-based grading, with the support of learning-for-mastery (Bloom, 1968, 1976) as the conceptual framework, to stage this research study, was confirmed through the literature and methodological reviews. The methodological approaches used for the research studies included in the literature review all sought to measure student achievement, motivation, and perception of student grades. The standards-based grading research is rooted in Marzano's theory of classroom assessment and standards-based grading (Heflebower et al., 2014). A mastery learning approach of assessment for learning instead of assessment of learning is primarily attributed to Bloom (1968). This theory provided the foundation for standards-based grading, with an emphasis on student outcomes. These theories provided the basis for the progression from assessment of learning to assessment for learning and grading practices that place a heavy weight and emphasis on summative assessment scores and outcome-driven measures in grade determination. The synthesis of these results provided a basis and justification for this research.

The evidence from researchers Pollio and Hochbein (2015) and Mitee and Obaitan (2015) clearly demonstrated that a mastery approach to instruction resulted in greater student gains. Standards-based grading has been shown to facilitate mastery learning because of the elongated assessment window with assessment serving as a tool for continued learning (Heflebower et al., 2014). Saefurrohman and Balinas (2016) evaluated the effectiveness of assessment-based grading and found that it provided an accurate measurement of student learning. Rosales (2013) indicated that a standards-based approach might not have an impact on student scores on an end-of-the-year exam. While this finding does not directly contradict a positive impact of a

standards-based approach found in other studies, it does indicate a need for further research focused on a comparative look at outcome-based grading practices like standards-based grading or assessment-based grading against a traditional total-points method.

Research indicated that student motivation increases with mastery-oriented approaches like standards-based and assessment-based grading. Brophy (2013) confirmed the findings through the lens of standards-based grading as conducive to student motivation and an increase in content mastery. These congruent findings provide assurances that grading practices rooted in mastery facilitate continued student learning and progress through increased student motivation. These studies confirmed Bloom's (1968) theory of mastery learning and established the continued relevance and value of mastery-oriented approaches in the classroom. It was clear that a mastery-approach with assessment for learning facilitated greater student learning. The purpose of assessment has shifted from a final measure of evaluation to a tool in the learning cycle to identify areas for continued student growth. This is a useful approach for instructors looking to maximize student learning.

The population of 56 Grade 8 math students used in this research study was a representative of a small rural school district with a high portion of students living in poverty (Midwestern Department of Education, 2017). Smeding et al. (2013) provided data that proved how a mastery approach can be used to help close the achievement gap between low and high SES students. Pekel (2013) provided evidence that assessment-based grades demonstrated a stronger correlation to state standardized math test scores than total-points grades for Grade 8 students. These results may prove to be important in the discussion of the implications of the research results from this study. Based on a cumulative fusion of the above findings, there are logical and sequential links that can be bridged to justify the warrant for this research. The

conceptual framework provided justification that legitimized the need for a research study to evaluate the effectiveness of assessment-based grading and to analyze the role the grading practice has in relation to mastery learning.

This research used a correlational methodological design similar to the study design of Hardegree (2012) and Rosales (2013) in the research review. This methodology was the most appropriate design to evaluate assessment-based grading in terms of the conceptual framework because compared the correlational relationship between classroom grades to standardized test scores. Hardegree (2012) demonstrated a statistical alignment between a standards-based grading approach and standardized test scores. Rosales (2013) measured the correlation between a standards-based approach versus a traditional total-points approach in comparison to posttest values. Both designs employed a quantitative approach to measuring the presence and strength of a correlation. Hardegree (2012) and Rosales (2013) served as a predicates for this research study, which sought to identify which of the two grading practices is a more accurate representative of student content knowledge. Overall, the literature review provided a conduit to better understand the issues related to grading and to better understand the purpose and role of assessment and grading, specifically using mastery learning as a mechanism for maximum student learning.

Critique of Previous Research

There are limiting aspects for applying the research results in a real-world setting. Even if a single grading strategy were proven to be the most effective, classroom grading will always be susceptible to the influence of instructor bias. Furthermore, the individual instructor would be accountable for using the tenets of the grading practice, especially for the policies rooted in the theory of mastery learning, using assessment as a tool in the feedback loop. Admittedly, this

individualized approach to instruction requires increased time and effort from instructors to provide a customized differentiated educational experience.

A recurring concern regarding the literature on a traditional total-points grading practice is the inconsistencies in grade determinations. Although a total-points approach is common in schools across the nation, the point assignments vary greatly and are left to the prerogative of individual teachers. A potential weakness of the study by Lee (2013) was shared with similar previous studies; there was no uniform directive on how the traditional total-points classroom grades should be determined. There was no mention in the study of the participating teachers being provided, or adhering to grading guidelines in regard to point-assignments for assignments and assessments (Lee, 2013). Ambiguity in the description of grading guidelines could lead to the assumption that each classroom teacher was allowed to determine point-assignments and types of activities included in the grade-book. Flexibility in grading means that classroom teachers may be using comprising traditional total-points grade differently (Wormeli, 2009). This variable potentially undermines the results of the study.

Some teachers hyperbolize assignment score determinations. Other teachers have the flexibility to weight assignment categories as they see fit. A teacher would have the opportunity to construct point-assignments to favor struggling students towards passing or to inflate grades. In an assessment-based grading practice, teachers are instructed to weight summative assessments to at least 80 % of the classroom grade, and there are other directives about types of activities that should be classified in the formative or summative categories (O'Connor, 2012). These recommendations help minimize the variance in the application of an assessment-based grading approach. These concerns were not addressed in the research included in the literature review. Welsh et al. (2013) found a moderate to weak correlation between standards-based

grades and test scores. This weak relationship may be explained by a weakness in the study. Further probing revealed that even in a standards-based grading setting some teachers still included noncognitive factors like participation and attendance, which all outcome-based grading practices should exclude according to the empirical review by Brookhart et al. (2016). The inclusion of nonacademic factors in any grading practices is a threat to the true validity of the grade or evaluated proficiency level (Brookhart et al., 2016).

A potential weakness of the study by Comes (2015) was the possible inaccuracies in the conversion of rubric-based CRG values into a comparable percentile grade. Hardegree (2012) had a potential weakness in the unknown accuracies of the reported standards achievements of students. Various instructors might define the mastery level of a standard differently. Student reports may also be influenced by teacher accommodations and modifications of student work and grades. Furthermore, there is the issue of potential inaccuracies faced by all standardized tests that cannot accurately measure all student skills and learning. A weakness of Fink's (2015) study was the lack of congruency in content; the two participating teachers had different content classes, one was an English teacher and the other taught math and science courses. It is possible that differences in student data were attributable to students' content strengths and weaknesses.

The current body of research had other limitations. The shared vulnerability in the critiqued literature is the potential for a common misunderstanding of what are standards-based practices. There is a frequent misuse of shared vocabulary between the terms standards-based and standards-referenced grading (Heflebower et al., 2014). This provides the potential for researchers to unintentionally assume they are using standards-based grading when in fact they are referencing standards-referenced grading practices. This could potentially skew the meaningfulness of the results if there was a misunderstanding on the researcher or instructor's

part. Research of this nature has to be conducted in a classroom setting, or at least provide a statistical analysis of classroom score and grade data. Rosales (2013) conducted research with an experimental design and correlational analysis but found a standards-based approach did not have a significant influence on student achievement as measured by a posttest when compared to a control group using a tradition total-points grading system. The results from Rosales (2013) did not find any relationship of statistical value contradict some of the other findings by Hardegree (2012) and provoked the need for further research. Fink (2015) included a relatively small study population, limiting the wide-scale application of the results. Pekel (2013) did not use cohort data, but did examine eighth-grade test scores across a 5-year period. Overall however, the body of research with consistent results and longevity of research-confirmed support for standards-based grading provided evidence-based assurances as to the effectiveness of standards-based grading practices.

To increase the validity of any study with a relatively small sample size, a larger more representative sample population would have provided a more comprehensive analysis, especially of sample sub-groups (Merki & Holmeier, 2015). Denson's (2013) research would have more reliable with a larger population size too; including 10th-grade students from other Florida schools would have further decreased the limitations of the study. As with similar studies, there was little background provided on individual teacher's total-points grading practices so, while the data was disaggregated by student demographics, the descriptive analysis did not include disaggregation by teacher which would have identified potential differences in classroom grading strategies.

Likewise, there were issues for consideration regarding the body of research on learning mastery. A large proportion of the research on the effectiveness of mastery learning by Mittee

and Obaitan (2015) and Smeding et al. (2013) took an analytic approach to already determined grades. Researchers' analysis evaluated reported grades and sought a statistically significant relationship between student grades and self-reported mastery-oriented instructional practices. Different instructors and students might define or recognize a mastery learning approach differently. Also, the large student populations that increase the confidence in the findings also make the study susceptible to increased influence from extenuating variables. There may be factors that affect student achievement, grades and performance that are not identified by the statistical analysis as explained by Brookhart et al. (2016). Despite this concern, the large population size of the studies by Beatty (2013), Knaack et al. (2012), and Smeding et al. (2013) increased the confidence in the applicability of the results on a widespread spectrum when applied in classrooms across the nation with diverse student populations.

The literature review attempted to mitigate the influence of outside variables associated with a large diverse study population (Rosales, 2013). The fact in this research study the grading method was not found to be an influential factor on student test score increases the need for further research (Rosales, 2013). The study by Mitee and Obaitan (2015) that did devise an experimental design to research with an experimental group had admittedly small student population numbers. Despite these limitations, there is a body of research on mastery learning since Bloom's (1968) that has provided evidence to confirm the effectiveness of mastery learning practices with high confidence.

Individual research projects by Knaack et al. (2012), Brophy (2013), Timmons (2017), Smeding et al. (2013), Pollio and Hochbein (2015), and Beatty (2013) have already widely established the effectiveness of mastery learning, standards-based learning; however less evidence exists to confirm or question the accuracy of assessment-based grading. Although the

sample size was small at 15 students and homogeneous in demographic makeup, the action-research study by Timmons (2017) was valuable because it evaluated the student perspective and receptiveness to being evaluated on mastery of content standards. Given what is known about the subject, based on the sources, it can be concluded that there is an important opportunity for this research not only to evaluate assessment-based grading practices. This research may also provide an evidence-based link to the accuracy of mastery-driven grading methods, or provide further support for total-points grading. Overall, the previous research studies were sound and with acknowledged merit. The research studies made logical claims that were rooted in theory and supported with evidence. The soundness of the previous research provides a background setting and basis for this research.

Summary of the Literature Review

The body of research related to grading demonstrated evidence regarding its central role in the field of education. The conceptual framework for this study was rooted in two major educational theories: Marzano, as cited by Heflebower et al. (2014), endorsed standards-based grading practice and proposal for mastery learning by Bloom (1968, 1974, 1976, 1984) as an instructional approach that supported the strategies in the grading system. Both of these theories view learning as a cumulative process with the end goal of maximizing student learning through a continuous feedback looping process. In addition to the body of research in support of standards-based grading, there was also research that specifically endorsed the use of homologous assessment-based grading practices. The pool of educational research specifically addressing the effectiveness of assessment-based grading was comparatively small. While this limited the discussion opportunities for a comprehensive literature review, it also highlighted the

opportunity and need for new research in this area to evaluate the scope and significance of effectiveness or inefficiencies of assessment-based grading strategies.

Given the information on the subject of student-mastery and the accuracy of academic grades, especially as they pertained to a standards-based evaluation, this literature review served as a summary of what educators and researchers currently know. Mastery learning has been proven to increase the overall learning; a larger body of content represented by mastery of learning objectives, for all students especially some subcategory groups like low-SES students (Smeding, et al., 2013). Research has shown that traditional total-points grading is an accurate representation of students' academic learning (Lee, 2013). Research has shown that assessment-based learning was an accurate representation of student learning (Saefurrohman & Balinas, 2016). Standards-based grading, a precursor to assessment-based grading, did not gain a foothold in public education until the early 2000's (Hamilton, Stecher, & Yuan, 2008). This presents a pivotal opportunity for this research study to evaluate the effectiveness of this trending grading issue in 2018.

A central question remains; does student grade achievement match tested achievement (Brookhart et al., 2016). There is a documented correlation between grades and test scores. After condensing a century of research, Brookhart et al. (2016) found that teacher-assigned grades typically correlate about 0.5 with standardized measures of student achievement. This value serves as a benchmark for the results from this research. Based on the review of literature, developed through a unique conceptual framework using standards-based grading and mastery learning to better understand assessment-based grading, there was sufficient reason for thinking that an investigation is warranted to examine the accuracy of assessment-based grading versus traditional-points grading in comparison to state standardized test values. Examining and

measuring the impact of two grading practices would yield socially significant findings. I can, therefore, claim that the literature review has provided strong support for pursuing this research project to answer the following multipart research question: What is the nature of assessment-based grading, how does a grading method influence the calculation of student class grades, and what relationship, if any, exists between classroom grades and standardized test scores? Next, in Chapter 3, the researcher will outline the methodology for this research study.

Chapter 3: Methodology

Mastery-directed grading practices, including standards-based and assessment-based grading, are becoming more commonplace in schools nationwide (Schinske & Tanner, 2014). Performance-based directives that require the academic grade be based solely on academic performance and knowledge have been gaining momentum and popularity, especially at the secondary level (Grinberg, 2014). More school districts are requiring teachers to use mastery-based grading practices (Spencer, 2017). There is a consistent emphasis, as is evident by the longevity and applicability of Bloom's (1968, 1974, 1976, 1984) principles of learning for mastery in the field of education, on practices that promote student mastery-learning (Adrian, 2012). The basic premise of assessment-based grading is that an academic grade should be primarily comprised of summative assessment scores (O'Connor & Wormeli, 2011).

Research has supported traditional total-points grading (Comes, 2015; Denson, 2013; Fink, 2015; Lee, 2013). Other research promoted standards-based grading and assessment-based grading to improve student achievement by prioritizing learning for mastery (Marzano & Heflebower, 2011). It was apparent from the literature in Chapter 2 that there is a gap in research; the research directly comparing how an academic grade would differ, using the same scores from a shared student population and configuring academic grades to compare scores using total-points versus assessment-based practices in comparison to a standardized test value, is lacking, specifically at the secondary level. Given the high-stakes emphasis placed on student achievement, school districts are under pressure to use a grading policy that best supports student learning (Supovitz, 2018). Statistical, empirical analytic data were needed to provide a research-supported recommendation to schools on a grading practice that best supports and promotes student learning. This study sought to determine a relationship between classroom grading

practices and students' academic grades in comparison to their standardized test score values. The results should be viewed through the lens of the theory of mastery learning, as it served as the basis for the assessment-based grading practice but is not included in the total-points grading.

Purpose of the Study

The literature review outlined the importance of selecting a grading procedure that is accurate in its representation of student learning. From the philosophical standpoint of a classroom teacher, one needs to be cognizant of choosing an instructional approach that fosters continued student learning. Many teachers want a grading policy that can achieve both. This research study sought to provide evidence towards identifying an effective grading practice at the secondary level. As outlined in the literature review in Chapter 2, different researchers advocated conflicting grading practices to achieve these goals (Iamarino, 2014; Knaack et al., 2012; Pekel, 2013; Saefurrohman & Balinas, 2016).

Researchers and practitioners who use a traditional grading model support the use of a total-points grading system as an accurate reflection of student learning (Comes, 2015; Denson, 2013; Fink, 2015; Lee, 2013). This researcher did not find research available specifically addressing if a total-points grading system better fosters a mastery-learning or continual learning approach. O'Connor and Wormeli (2011) disagreed with the accuracy of a total-points grading approach and used evidence to support a performance-based grading approach such as standards-based and assessment-based grading. Furthermore, standards-based and assessment-based approaches used Bloom's (1968) learning for mastery pillars and were supported by researchers Saefurrohman and Balinas (2016), Mitee and Obaitan (2015), Smeding et al. (2013), Pekel (2013), Pollio and Hochbein (2015), and Beatty (2013). Previous research results by Iamarino (2014) individually isolated components of the dynamic between accuracy in grading in a

traditional versus an assessment-based practice. Brophy (2013) conducted research that provided results that promoted the use of a grading practice that incorporated the tenets of learning for mastery as described by Bloom (1968) through multiple assessment opportunities. The research gap indicated a need to simultaneously compare the two grading practices against a standard value using a single student population. The research study addressed this gap in current research.

The purpose of this study was to determine whether a relationship exists between the grading practices used by classroom mathematics teachers of a small rural school and student performance, as an indication of summative learning, on the math MAP test, a standardized content assessment. It sought to identify the strength of any relationship identified by the evidence-based results, as they may be applicable for future decisions on grading practices for this school. The study also sought to determine which grading practices are in better alignment with learning-for-mastery standards and result in greater overall student gains. In interpreting the results, this study could play a key role in bridging the gap to identify a relationship between student performance evaluated on a standardized test and one of the two described classroom-grading policies. This information may be helpful in providing evidence-based recommendations of grading practices as an application of the results from this rural school. The purpose of this quantitative research study was to determine if and where a difference existed in the academic percentile grades for eighth-grade math students in a regular education math classroom when calculated using two different grading practices. Students' grades were determined using a total-points and an assessment-based approach, in comparison to standardized test scores, controlling for student assignment scores for eighth-grade middle school math students at a small rural school.

Research Questions

The main objective of this study was to explore how student classroom grades may differ if calculated using different grading practices. This researcher accomplished this by calculating and comparing academic grades for each Grade 8 math student using two different grading practices: total-points grading and assessment-based grading. The researcher also sought to identify the presence or strength of the relationship between each student's classroom grades and his or her math MAP standardized test scores. The central research question guided this study: Which grading method determines classroom grades for Grade 8 math students that most accurately align to corresponding standardized math MAP test score values for Grade 8 math students? Under this central research question, the specific research questions addressed in this study included the following:

1. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students?
2. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students?
3. What was the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students?
4. Which grading method (traditional total-points grading and assessment-based grading) has resultant scores that more closely aligned to students' standardized test scores?

Hypotheses

With the research question: What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students?

Directional hypothesis: There was a statistically significant difference between mean values of the semester classroom grades derived from two different grading methods, traditional total-points grading and assessment-based grading, for Grade 8 math students.

Null hypothesis: There was no difference between mean values of the semester classroom grades derived from two different grading methods, traditional total-points grading and assessment-based grading, for Grade 8 math students.

With the research question: What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students?

Directional hypothesis1: There was a statistically significant difference between mean values of the semester classroom grades derived from traditional total-points grading and MAP standardized test score for Grade 8 math students.

Null hypothesis1: There was no difference between mean values of the semester classroom grades derived from traditional total-points grading and MAP standardized test score for Grade 8 math students.

Directional hypothesis2: There was a statistically significant difference between mean values of the semester classroom grades derived from assessment-based grading and MAP standardized test score for Grade 8 math students.

Null hypothesis2: There was no difference between mean values of the semester classroom grades derived from assessment-based grading and MAP standardized test score for Grade 8 math students.

With the research question: What was the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students?

Directional hypothesis1: There was a positive relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students.

Null hypothesis1: There was no relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students.

Directional hypothesis2: There was a positive relationship between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.

Null hypothesis2: There was no relationship between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.

With the research question: Which grading method (traditional total-points grading and assessment-based grading) had resultant scores that more closely aligned to students' standardized test scores?

Directional hypothesis: There was a stronger positive relationship between the grades derived from assessment-based grading practices, standardized test score values for eighth-grade math students than between the grades derived from traditional total-points grading, and standardized test score values for eighth-grade math students.

Null hypothesis: There was no difference in the relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students and between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.

A directional hypothesis for this research question is rooted in the assumption that assessment-based learning is more conducive to mastery learning because of the feedback loop with multiple assessments, additional teaching opportunities and opportunities for repeated assessment opportunities as outlined by Bloom (1968). Traditional total-points grading did not offer summative assessment retake opportunities and therefor did not incorporate the tenets of a learning-for-mastery approach. Standardized test scores should measure each student's cumulative content knowledge, to provide an evaluation of the content that the student has mastered. Classroom summative assessments should also measure cumulative content knowledge. The null hypothesis played an important role in the logic of statistical testing; however, the proposed directional hypothesis sought to provide a clear link between the conceptual framework and the research questions.

Research Design

The methodological design best suited for this research, based on the purpose and research questions in this study, was a quantitative approach with a correlational, non-experimental research design. This design was most appropriate because the researcher was looking for an association between variables after an event, the instruction and assignment grading, has taken place. Denson (2013) used a similar approach with a correlational, non-experimental research design to examine the relationship between teacher-assigned grades and students' test scores received at state's standardized test, Florida Comprehensive Assessment

Test, overall and controlled for student characteristics like gender and ethnicity. Lee (2013) also used a similar correlational mixed-methods approach to address a research question centered on the relationship between teacher-assigned grades and state standardized test scores, as measured by the Arizona Instrument to Measure Standards (AIMS). This was the most appropriate methodology selection for researchers Denson (2013) and Lee (2013) and likewise was the best research study design choice for this research study.

This study took the set of raw scores from students' classroom assignments and assessments and applied two different grading strategies, a traditional total-points practice versus assessment-based practice, to evaluate the relationship between grading policy and academic grades. It can be assumed that the total-points grading practice does not specifically include strategies of Bloom's (1974) mastery learning because it does not include reteaching and reassessment to increase students' gains in mastery. The researcher was not able to randomly assign the students into control and treatment groups because the study used archival data where the instruction has already occurred by another teacher. First, the researcher used a paired samples t-test to determine what difference, if any, existed between students' total-points classroom semester grades and assessment-based classroom semester grades without test retake opportunities that were calculated using the same set of assignment and assessment scores for Grade 8 math students. This analysis was important because it could isolate the impact, if any, resulting from the weighted summative and formative categories that are required in the assessment-based grading practice.

A second paired samples t-test analysis was used to determine what difference, if any, existed between students' total-points classroom semester grades and assessment-based classroom semester grades with updated test scores reflecting any higher test retake scores; the

assessment-based grading classroom semester grades was still be calculated using the 90% summative 10% formative weighted categories. This analysis isolated the direct influence of mastery-learning practices, as was evidenced through improved retake assessment scores, on a classroom grade. Furthermore, a similar analysis compared grades from each grading practice against a standardized MAP math test score. The corresponding academic total-points and assessment-based grades were compared through a paired samples t-test to the MAP math standardized test scores to determine which resulting grade was a better representation of students' content knowledge.

Correlations. The next stage of the analysis sought to investigate a correlational relationship between grades derived from each grading practice and a standardized test score. The researcher calculated the Pearson correlation coefficient to determine whether a positive relationship existed between grading practices and MAP standardized test scores. Then, the correlation coefficients were compared to determine whether the relationship between assessment-based grades and MAP standardized test scores was stronger than the relationship between total-points classroom semester grades and MAP standardized test scores. The analysis results and discussion work to identify the class grading practice that more closely aligns to the standardized score with statistical significance. To achieve this statistical analysis, the researcher needed individual score data from students, basic demographic data to code students, and standardized test scores. Any researcher, if avoidable, should not contribute to the data collected for analysis to increase the validity and integrity of the study. This minimizes the influence of researcher bias by protecting the data from contamination as a result of researcher bias (Pannucci & Wilkins, 2010).

This research study used data in the form of student assignment scores, assessment scores, and standardized test scores from eighth-grade mathematics students. In a small rural school, a single teacher often has an entire grade of students in his or her classes; this mitigated the influence of instructional differences among different teachers in score outcomes. This researcher used previous scores from an entire second semester of student scores from all regular education eighth-grade math students. Correlation methodology was the most appropriate research design selection because used the data that already existed to identify a relationship between the grading practices after the instruction and standardized testing had already occurred.

Description of study population, sampling method, and study procedures. This study included participants from one secondary mathematics teacher, referred to with the pseudonym Mrs. Smith, with over 30 years of teaching experience in math education. Mrs. Smith had four sections of regular-education eighth-grade Pre-Algebra students in a small rural school in the Midwest. The district served approximately 400 students total in grades 7–12, with a special education rate of about 24% and about 50.33% of the students qualify for free and reduced lunches (Midwestern Department of Education, 2017). There was minimal ethnic and racial diversity with less than 2% of the student population classified as English language learners, 90% of the students reported being white and non-Hispanic, and the graduation rate was 93.44% for 2017 (Midwestern Department of Education, 2017).

No convenience sample was used for the study; the study population consisted of 57 students, including all eighth-graders in the regular education math classroom. As a result of the class registration, the target population of all regular education eighth-grade math students had the same experienced instructor, Mrs. Smith. These parameters dictated a purposive sampling approach was most appropriate for the outlined research design. An inclusive selection approach

reduced opportunities for researcher bias in the definition and selection of the sampling frame and maximized the population size for this study. The target population was large enough to produce meaningful data but small enough that a subset sampling method was not necessary. The data included assignment and assessment scores for each participant in the population for a semester, or 18-week grading term.

Instrumentation

For the instrumentation for data collection, this researcher accessed previously determined student grades and test scores from Canvas, an online learning management system used by the local school district, and PowerSchool, the online grade book. The math curriculum series used by the host district was Saxon Math, which is a comprehensive curriculum that includes daily lessons built into units and summative assessments (Houghton Mufflin Harcourt, 2019). For summative classroom assessments, Saxon assessments that corresponded to each unit of lessons were administered. The classroom assessments were not multiple choice; students were given formula sheets and expected to calculate the correct numerical answers (Houghton Mifflin Harcourt, 2019). The Saxon Math series was aligned with the national Common Core math standards (Beltzner, 2018). The NWEA MAP math assessment was also aligned to the national Common Core math standards (Set, 2018). This was important to ensure that the classroom assignments and assessments taught and evaluated the same Common Core math standards and learning objectives as the MAP standardized test to make the correlational analysis possible. The MAP test was multiple choice, students were expected to calculate the correct answer and select it from four answer choices (NWEA, 2018). Both classroom Saxon assessments and MAP standardized assessment were administered in the typical classroom setting by the instructing math teacher.

Basic demographic data for students including gender, ethnicity and socioeconomic, or SES status, was also available for coding purposes but not used as a category for analysis. This readily available archival data provided all of the information needed to conduct this research study to answer the research questions. The grading analysis was used to identify a causative relationship through a correlation analysis of the two sets of class grades determined using different grading practices. It was important to note that a correlational study design worked to identify a relationship between the variables but cannot provide a cause-and-effect certainty.

Data collection. The researcher used previous student scores as a data source. There was not a convenience sample for the study because the study body size was relatively small. The study sample size consisted of 56 of the 58 eighth-grade students in the general mathematics class. The study potentially included 58 students; however, one student was eliminated from the study due to a truant status. Another student enrolled in the district with less than half of the semester remaining and was excluded from the study as that student had missed more than half of the instruction from the participating teacher. The final study population included 56 students of the potential 58 student contributing to the data set. Other students excluded from the study population included students who were at a math ability level of at least three grade levels behind and were in a remedial special education math resources class in compliance with their individual IEPs instead of in a general math class. All students were taught by the same instructor, who is a seasoned secondary math teacher with over 30-years of teaching experience in math education.

The participating classroom teacher and district administrative official provided written approval to serve as evidence of informed consent in support of ethical research standards prior to any data collection and analysis. The researcher, through a coding process, protected student

names and demographics and the statistical analysis was summarized as a whole class in the results. Student assignment and assessment scores were accessed through the grades archives in the PowerSchool online grade book system used by the district. The participating instructor proctored classroom assessments to the students. The classroom assessments were included in the Saxon curriculum and textbook materials used for the class (Houghton Mifflin Harcourt, 2018). The Saxon Math curriculum series was used from kindergarten through 12th grade at the participating school district. The use of assessments included in the curriculum series ensured that Mrs. Smith's grading methods were valid and sound and provided an accurate evaluation of students' learning in accordance with the learning objectives. Research has shown that the Saxon Math program has demonstrated instructional effectiveness by increasing students' knowledge and understanding of mathematics (Educational Research Institute of America, 2009).

The standardized test scores that were used came from the administration of the Measures of Academic Progress, or MAP test, a nationally recognized assessment that measures the performance of every student to evaluate if every student is achieving at, above, or below grade level for the specified content ("MAP," 2018). The mathematics MAP assessment is an adaptive test because the question difficulty level continues to change based on each student's performance. The MAP assessment is an evaluation tool to provide accurate personalized assessment of student content and tracks growth over time ("MAP," 2018). The MAP test was aligned to both the state standards from the Midwestern Department of Education and the nationally recognized Common Core mathematics standards and uses anonymous data from over 10.2 million students to create national norms ("MAP," 2018). These measures made the MAP

mathematics test a sound selection as the source for standardized test scores for students and served as a benchmark for the grade comparison outlined in Chapter 4.

Operationalization of Variables

It was important, especially in a quantitative study, to define variables in measurable terms. In this research study, the term academic grade or grade referred to the percentile value of the total points earned divided by the total points available converted to a decimal and assigned the corresponding letter grade. The academic grade was the semester grade, which was an average of the two quarterly grades. The term score meant the points earned divided by the points available converted into a percentile value for an individual assignment or assessment.

The participating school district that provided the setting for this research study, similar to the other schools in the regional conference, had a school-wide grading scale where the following grades were assigned (XXXXX & XXXXXXXX, 2017). That grading system is displayed in Table 1.

Table 1

Letter Grades by Percentile Score

Percentile score	Letter grade
93%–100%	A
85%–92%	B
77%–84%	C
70%–76%	D
0%–69%	F

Any reference to a standardized test score for this study was referring to a RIT, or Rasch Unit, which was the students' score values produced by the math MAP test. For this study the RIT score was be converted into a percentile score. A grading practice was defined by a policy that dictated how academic grades were calculated using a raw total points approach or some combination of total points within designated weighted categories. The assessment-based

grading practice, as described by O'Connor (2012), has a summative weighted category that constitutes at least 90% of the overall class academic grade.

The study population had clear parameters established through enrollment in the public-school system that already categorized students by age groups into grade bands. All eighth-grade students in the district enrolled in regular education math were included in the study population. Study participants were excluded if the attendance record showed that the individual missed 10 or more days cumulative for the semester as consequently the student would have missed a substantial portion of the class instruction. The Midwesternstate's Compulsory Attendance Statute XX-XXX mandates that students shall not be absent more than 20 days per year. If this occurs, this school district pursues truancy actions. Students are not permitted to miss more than 10 days per semester to avoid truancy proceedings. In this study, any student that was absent more than ten days per semester was considered a truant student and as not be included in the study's student population.

Data Analysis

The researcher was granted research approval for this study from the Institute Review Board (IRB) of Concordia University–Portland prior to all data collection. Upon completion of all coursework and standardized testing, the district's data steward provided the researcher with a score sheet printout for each student with scores from class assignments and assessments, and standardized MAP math test scores for the spring 2018 semester. The researcher imported the data into the IBM SPSS Statistics software for the purpose of analysis. For the first and second research questions, the variables representing academic grades based on two grading practices and standardized test scores were measured on interval ratio, and therefore could be compared using paired samples t-test as the grades were received for the same students.

To answer the third research question, Pearson correlation was used to determine any relationship between academic grades derived from the traditional total-points grading practice and the assessment-based grading practice and standardized MAP math test. For the last research question, to investigate the difference between the strength of the relationship between students' grades derived from each of the grading practices and standardized test scores, the researcher used a special free statistical tool (Diedenhofen & Musch, 2015) to implement a correlation coefficient comparison procedure first described by Hittner, May & Silver (2003). To determine if a relationship was statistically significant, the researcher used a threshold confidence level of $\alpha = .05$ for each evaluation. Furthermore, any results with a *p*-value of less than .05 were viewed as significant and should be considered for further or continued research.

Limitations and Delimitations of the Research Design

A limitation for this study included the selection of the population for this study. Eighth-grade students in special education math courses as directed by Individual Education Plans, or IEPs, were excluded from the study because they do not receive the same instruction, assignments and assessments as their regular education peers. The researcher has no control over students' math class registration. The school's guidance counselor created student schedules. There was only one regular education Grade 8 math teacher; all eighth-grade regular education math students were enrolled in the participating teacher's math class. The researcher used score and grade data from another teacher so student scores were not self-reported and the researcher did not contribute to the data.

While all students were given the same assessments by the participating teacher, Mrs. Smith (pseudonym) did inherently have some influence in how classroom grades were determined. The risks to this were minimal as student Saxon assignments and assessments were

identical and scored for correctness. Any potential limitation was centralized in the possibility of the participating teacher awarding partial credit for incorrect student work or answers. It is assumed that Mrs. Smith was accurate and consistent in determining the points assigned on assessment problems across the student population. NWEA scored all of the MAP math tests for the student population, so the MAP tests were not scored by the participating teacher. Another limitation was that not every student chose to do a test retake to improve their summative assessment scores in the assessment-based grading method. This researcher is not able to predict how student scores may have been affected by the available retake opportunity and the data analysis was conducted on the actual recorded scores. The researcher previously taught most of the same student population in science class during the 2016–2017 school year, the academic year before this study was conducted and in a different subject area.

Delimitations for this study included the parameters for selecting student assessment scores and the time frame from which data were accessed. In a traditional total-points grading practice, test retake opportunities for a higher replacement grade were usually not provided so the class grade calculations when determining the academic grades with scores using a total-points method consistently used the first recorded assessment score on summative assessments. Assessment-based grading used principles of Bloom's (1968) learning for mastery and used the highest score, including retake opportunities, in the class grade calculation. The same assignment scores values were used in both grading practices, but were weighted differently categorically in correspondence with the prescribed grading practice. Instrumentation included accessing previously recorded data in the form of student scores and grades from another experienced teacher in the same building, but not the researcher's students, to mitigate threats to the validity of the study's findings. Furthermore, the term of one semester was determined as an

appropriate length of time to provide enough data to be statistically significant while remaining manageable for the population size. The second semester of the academic school year, or spring semester in a nine-month school year, was selected so that the standardized test scores would reflect cumulative student content learning. The researcher acknowledged these limitations and delimitations in an effort to provide transparency in the study and reduce potential threats to validity.

Internal and External Validity

The quantitative correlational research method used the statistical analysis of previously recorded data to provide evidence of the absence or presence and strength of a relationship between variables. Internal validity is best achieved through results that are measurable. Two acknowledged weaknesses in any correlational research design are the inability to manipulate the independent variable and lack of randomization. Because the researcher could not assign student groups due to registration policies, there was no randomization or sampling method. To address this concern the researcher expanded the target population to include all eighth-grade regular education math students instead of creating any sub groups or sampling to relieve the possibility of subject selection bias. The researcher was also unable to directly manipulate the independent variable because the study methodology was not an experimental design. However, using existing data from an outside source to which the researcher did not contribute or have the ability to influence mitigated concerns of a threat to internal validity of this study.

The researcher of this study acknowledged concerns with external validity. The lack of ethnic and racial diversity in the study population, due to low demographic diversity in the school district setting, may influence the widespread applicability of the research findings. Any strong statistical relationship identified between variables in the data analysis should be

generalized and considered applicable only to a comparable audience. Nevertheless, the results may add to the current body of literature on topic of grading and warrant further study with a similar research design using a different, expanded diverse study population.

Expected Findings

Grading will inevitably remain a controversial issue in education. Not all students can be evaluated and quantified in the same manner. In this study, this researcher used the same assignment score values and assessment score values for each student, but evaluated two grading practices and the impact of an extended timeline for learning the same material. Not every grading policy or practice promotes continued learning or learning for mastery. Teachers and stakeholders want to assign grades that are accurate and fair and should be looking for evidence-based recommendations to help influence grading policy decisions. This study sought to provide statistically derived information on grading practices that can be applied by classroom teachers and administrators. From the results of this study, this researcher looked to interpret the relationship between classroom grades and standardized test scores. Through this study, the researcher hopes to better understand how to use grading to promote continued student learning towards mastery.

Ethical Issues in the Study

This study met all ethical standards for beneficence and nonmaleficence because student and teacher rights and welfare were protected. The cooperating teacher in the research study voluntarily offered the researcher access to student scores, grading, and demographic data with approval from the administration so the researcher could maintain professional boundaries and responsibilities. The researcher upheld the highest levels of integrity maintaining transparency in records and reporting of accurate findings. The researcher respected student and staff privacy

and confidentiality by using safeguards that protected any identifiable information and worked to be aware and respectful of cultural and role differences. Participants were not given the opportunity to volunteer to participate, so it was vital to protect individual identities and use anonymous data with descriptors. The researcher, as a classroom instructor, had personally used both grading practices discussed, a total-points method and assessment-based grading, and saw value, advantages and disadvantages in each. The researcher conducted an accurate and honest statistical assessment independent of any personal or philosophical ideals about the preference in classroom grading practices. The researcher collected information that included student names and scores but did not use any identifying information to protect the privacy of students and ensure confidentiality. All demographic and identifying data were coded will be hosted on a protected server for a period of five years before being permanently deleted.

Summary

This chapter has outlined the research design of this study including strategies for instrumentation, data analysis and selection of participants. The researcher worked to recruit a participating teacher that granted access to the necessary student data and generated a population size large enough to produce meaningful analysis. The research questions could not be answered without a quantitative approach because the difference in grade mean values could not be calculated using a qualitative analysis. The utilization of previously recorded scores and grades made a correlational analysis the most appropriate design choice. This methodology choice helped identify what, if any, relationship existed between grades and standardized test scores. Instrumentation included access to previously determined archived student scores and applied different grading practices, traditional total-points and assessment-based approaches. A correlation analysis provided feedback on the strength of any relationship between the grading

variables. A correlational methodology was the best selection to answer the research questions with the available data within the confines of instrumentation while using the largest sample size available. Next, Chapter 4 presents and summarizes the results and findings from this research study.

Chapter 4: Data Analysis and Results

Introduction

This chapter presents the results of the statistical analysis performed on the students' grades that were based on different grading practices: traditional total-points grading, assessment-based grading and compared to math MAP standardized test scores. The total-points grading practice and an assessment-based grading practice used the same set of student scores from the study population but weighted them to calculate the final class grades in different ways. The researcher also ran an analysis to isolate the impact of test retake scores, as it served as the functional application of Bloom's (1976) theory of mastery learning. The purpose of this study was to provide an analysis that sought to identify correlations, if any, between grades derived from different grading practices using the same student scores. These values were compared to Grade 8 math student standardized MAP test scores to see which grading practice more closely aligned to standardized test scores. The difference between mean scores and the relationship between the three grade measures were evaluated. The analysis followed the research questions and corresponding hypotheses presented in Chapter 3.

The general theme of inquiry on which the data analysis was based was to establish the average difference in means between the sets of data, specifically as each set compared to the students' standardized MAP test scores. The study design was a correlational quantitative methodology to identify any relationships that already existed between groups of grading data. This was the most appropriate choice because the data were archived and there were more than two groups of variables. Scores were retrieved for 56 Grade 8 math students from a small rural school in the Midwest. There was no subsampling; all students enrolled under the participating teacher for the semester who were not classified as truant were included in this correlational

study. The main limitation of the study was the low ethnic and cultural diversity of the study population due to local demographics and enrollment. Delimitations were addressed by using a large enough study population and a participating teacher so that the researcher did not add to or influence the achieved data retrieved for this study. Instrumentation strategies stemmed from the correlational methodology and included paired samples t-test and Pearson correlation analysis, and a Shapiro-Wilk test to evaluate normality and minimize any instrumentation threat to validity and increase reliability of results. In Chapter 4, the researcher will describe the study sample and population, provide a summary of the analysis results, and a detailed analysis of the grading data in relation to each research question.

Description of the Sample

The study population included 56 Grade 8 math students enrolled in the participating teacher's pseudonym, Mrs. Smith's math classes for the spring 2018 semester. All semester grades were calculated as percentages to two decimal places. Student math MAP scores were reported as a nationally-norm referenced value and percentile score. Student percentile scores were not useful as percentile scores in comparing student grade means through correlational analysis, as it was a different method from the grade calculation approach used with classroom grading. This disjunction could potentially have skewed the data analysis results. To ensure the most accurate analysis, the RIT score needed to be converted into a percentage. NWEA provided a spring MAP math score norm that represents grade level learning as 230.9 with a standard deviation of 19.11 (Summit Learning, 2015) as shown in Figure 3 in Appendix B. Classroom grades derived from traditional total-points grading practices have also shown to fall under a bell curve, as shown in Figure 7 in Appendix B, with each standard deviation a change in letter grade as first presented by Bloom (1976) and explained by Guskey (2005). This was

helpful because it allowed the researcher to align the benchmarks provided by each set of MAP and letter grade standard deviations to create a conversion scale as outlined in Table 7 in Appendix B. This conversion chart provided the key to converting students' MAP RIT scores to a percentage, but did not provide the precision of two decimal places to match the number format of the grades. The researcher knows there is a positive relationship between the RIT and percentage, and using the scale conversion can calculate the slope to be 0.5233. The researcher was able to use the mathematical equation for the point-point formula of a line to find the equation of the conversion to more precisely calculate the RIT to percentage conversion, as shown in Figure 2. This equation was utilized on the grading spreadsheet. The conversion equation provided a better solution to the challenge with student RIT scores and allowed the researcher to accurately conduct the correlational analysis.

MAP RIT to Percentage Conversion Equation
MAP year-end mean = 230.9, SD = +- 19.11
y = 0.5233x - 40.83
y = percentage, x = RIT score

Figure 3. This figure shows the mathematical formula for the conversion of a student's MAP RIT score into a percentage based on the point-point mathematical formula for line. The x-value represents the RIT score input and the y-value represents the student's percentage for the standardized test.

Table 2 presents the mean values of the students' scores for each measure along with the standard deviation. All variables were checked for normality of the distribution using the Shapiro-Wilk test. These results are also included in Table 2.

Table 2

Students' Scores Descriptive Statistics and Normality Tests Results

	M	SD	Shapiro-Wilk test	Shapiro-Wilk *p*-value (*df* = 56)
Total-points	82.7	7.4	.969	.166
Assessment-based	81.4	7.8	.974	.259
Assessment-based Retakes	82.4	7.3	.976	.321
MAP Scores	91.7	5.9	.944	.012

All mean math scores of Grade 8 students were relatively high (> 80). Students received a mean score of 82.7 (*SD* = 7.4) based on traditional total points grading. The mean score received at assessment-based grading was slightly lower (*M* = 81.4, *SD* = 7.8) but was higher with retakes (*M* = 82.4, *SD* = 7.3).

Both paired samples t-test and Pearson correlation analysis require that the distribution of the variables is approximately normal. Shapiro-Wilk test was conducted to test normality. The results showed that total points and both measures of assessment-based grades, before and after retake, are normally distributed (Shapiro-Wilk p-values are greater than 0.5). The distribution of MAP standardized test scores did not fit normal distribution properties based on Shapiro-Wilk test ($p = 0.012$). However, taking into account the skewness and kurtosis of students' standardized test scores lie within the borderline of -1 to 1 (skewness was equal to -0.594 and kurtosis to -0.504), the distribution can be assumed as approximately normal and therefore both analyses techniques, t-test and Pearson correlation can be used.

Summary of the Results

The researcher took many steps to ensure the accuracy and precision of results from the statistical analysis. A contracted data specialist externally audited all score entries, grade calculations, and statistical correlational analysis for accuracies. This ensured the precision of the calculated grades and analysis results. Percentages were reported as a number and rounded

to two decimal places. Spreadsheet calculations used formulas where appropriate to increase the precision and accuracy in grade calculations. The grading spreadsheet was password-protected and housed on a secure network. The summary of grades is presented in the Appendix A. The following detailed analysis outlines how total-points grades compared to assessment-based grades with and without retake scores and evaluated the relationship of each classroom grade to student math MAP scores.

Detailed Analysis. For each research question, a corresponding hypothesis was formulated in Chapter 3. The detailed analysis for each research, directional and null hypotheses and statistical analysis is outlined in the subsequent sections.

Difference between grades derived from different grading methods. The students' assessment-based scores were measured twice: before and after retake, therefore every hypothesis (see Table 3) including assessment-based grades was checked twice: first for scores received before and then after retake.

Table 3

Hypotheses for Research Question 1

Research question 1	Directional hypothesis	Null hypothesis
What is the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students?	There is a statistically significant difference between mean values of the semester classroom grades derived from two different grading methods, traditional total-points grading and assessment-based grading, for Grade 8 math students.	There is no difference between mean values of the semester classroom grades derived from two different grading methods, traditional total-points grading and assessment-based grading, for Grade 8 math students.

To answer this research question, the mean grades derived from a total-points grading practice and the two measures of assessment-based grades (before and after retake) were compared using paired samples t-test. This assessed whether the difference between the mean values was

statistically significantly different from zero. The results of the test, outlined in Table 4, showed that traditional total-points were significantly higher than those received through the assessment-based grading method (mean difference of 1.3 points was statistically significant as proved by t (55) = 7.31, $p < 0.001$). While the assessment-based scores received after retakes did not show a statistically significant difference from total-points derived grades. These results illustrated that there was a difference between grades derived from total points grading method and assessment-based grades before, but not after retakes.

Table 4

Students' Grades Difference by Grading Method

	Mean difference	*SD*	*t* (*df* = 55)	*p*-value
Total Points - Assessment-Based	1.3	1.4	7.31	< 0.001
Total Points - Assessment-Based Retakes	0.3	1.8	1.15	0.255
Total Points - MAP Scores	-9.0	7.1	-9.56	< 0.001
Assessment-Based- MAP Scores	-10.4	7.7	-10.14	< 0.001
Assessment-Based Retakes- MAP Scores	-9.3	7.4	-9.36	< 0.001

Difference in grades compared to test scores. The results showed (Table 2) that all grades received from both traditional total scores and both assessment-based methods (before and after retake) differed significantly compared to MAP standardized test score for Grade 8 math students. The greatest difference of -10.4 was between assessment-based grades without a retake and standardized tests score (t (55) = -10.14, $p < 0.001$), while the difference between assessment-based grades with retakes compared with test scores was lower -9.3, but also highly statistically significant (t (55) = -9.36, $p < 0.001$). The least difference but also highly significant was between traditional total-points grades and standardized test scores (M = -9.0, t (55) = -9.56, $p < 0.001$). Taking into account that there was a statistically significant difference between total-points and assessment-based grades before retakes, it can be concluded that the difference between total-points calculated grades and standardized test scores was significantly lower than

the difference between assessment-based scores and standardized test scores. These results answer the second research question (see Table 5) showing significant lower grades for all grading methods compared to standardized MAP test scores with the smallest difference obtained from total-points grades.

Table 5

Hypotheses for Research Question 2

Research question 2	Directional hypothesis	Null hypothesis
What is the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students?	There is a statistically significant difference between mean values of the semester classroom grades derived from traditional total-points grading and MAP standardized test score for Grade 8 math students. There is a statistically significant difference between mean values of the semester classroom grades derived from assessment-based grading and MAP standardized test score for Grade 8 math students.	There is no difference between mean values of the semester classroom grades derived from traditional total-points grading and MAP standardized test score for Grade 8 math students. There is no difference between mean values of the semester classroom grades derived from assessment-based grading and MAP standardized test score for Grade 8 math students.

Relationship between grades and test scores. First, a set of scatterplots shown in Figure 4 and 5, were visually inspected to find out whether there was a linear relationship between the variables so that a Pearson correlation could be used to estimate the strength and significance of the relationship. All scatterplots showed that there as a positive linear relationship between grades of different grading method and students' standardized MAP math test scores. These visuals provided as evidence in support of the directional hypothesis for research question 3 as outlined in Table 6.

Table 6

Hypotheses for Research Question 3

Research question 3	Directional hypothesis	Null hypothesis
What is the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students?	There is a positive relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students.	There is no relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students.
	There is a positive relationship between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.	There is no relationship between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.

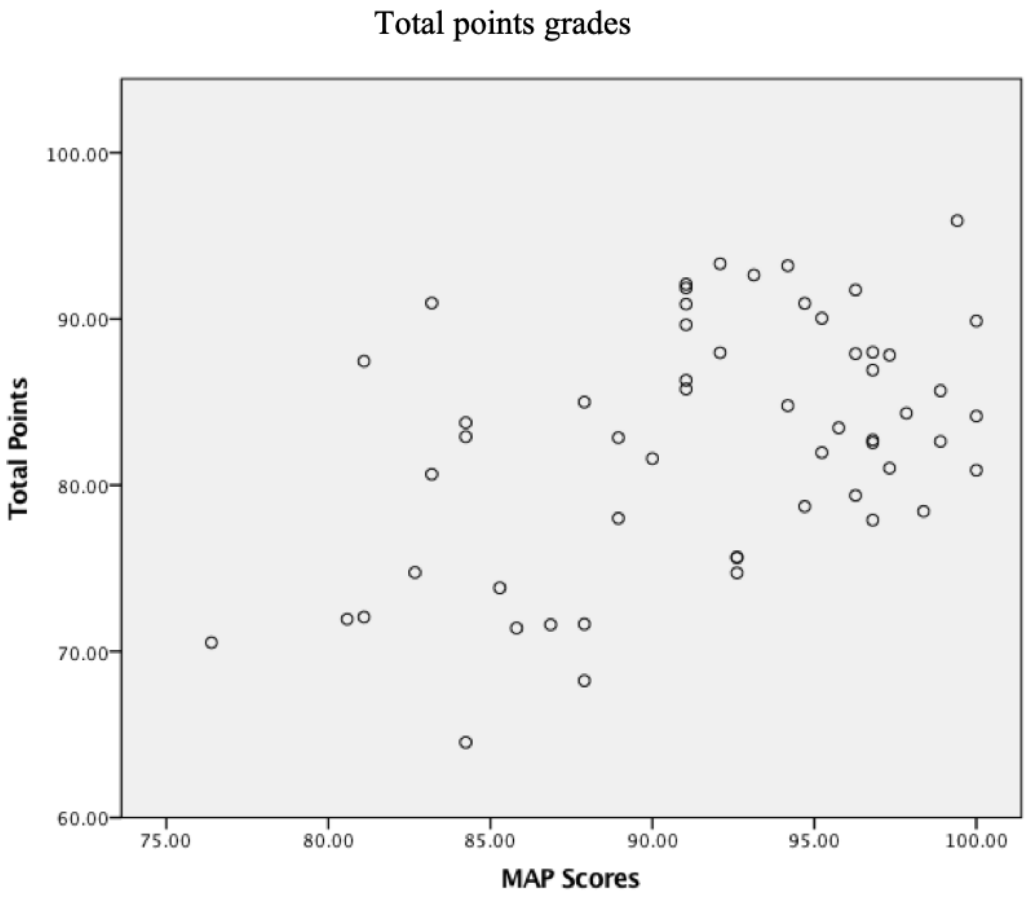

Figure 4. Total-points grade scatterplot by MAP math test scores.

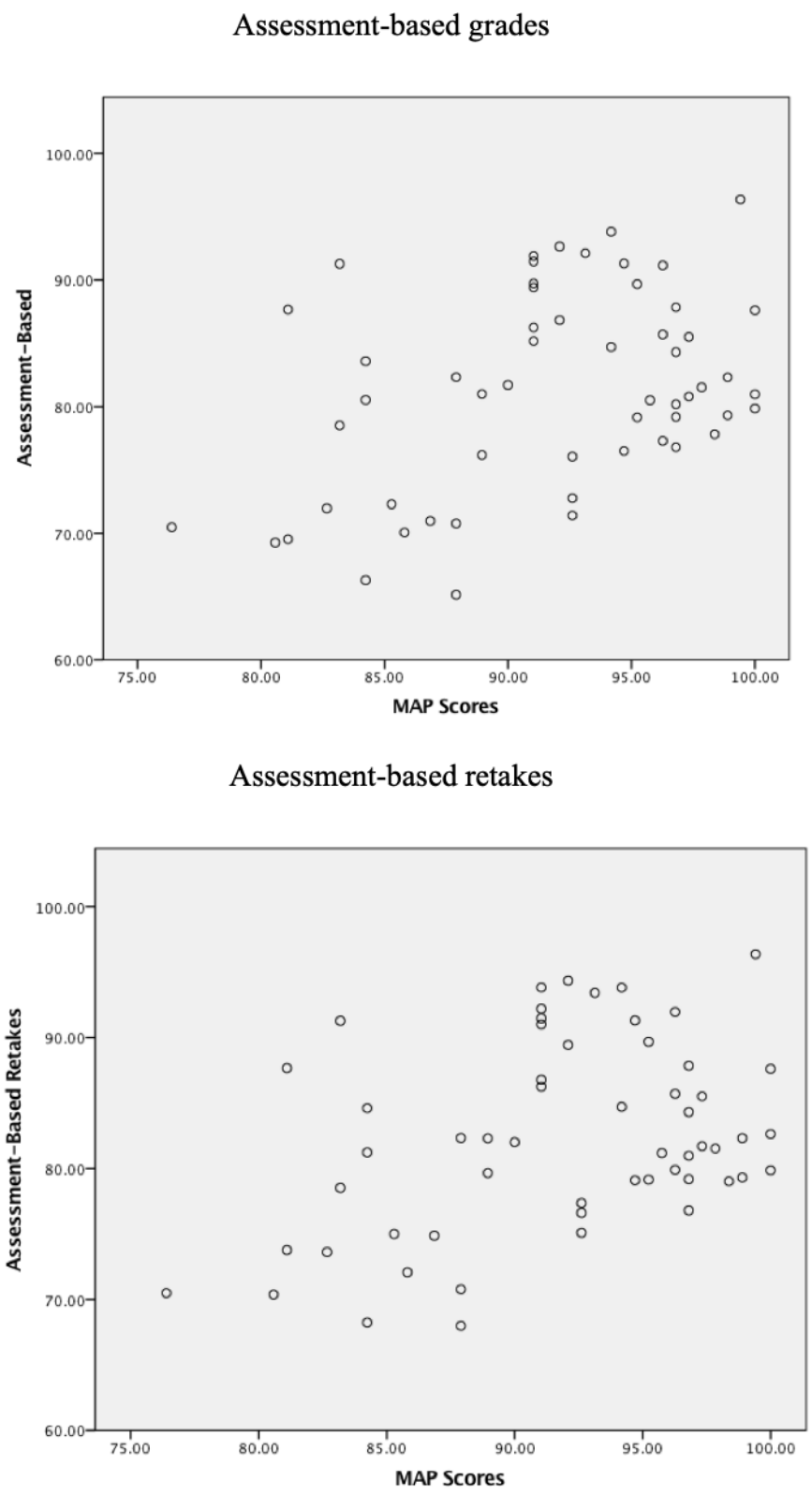

Figure 5. Assessment-based grades scatterplots by MAP math test scores.

A Pearson correlation analysis was performed to evaluate the relationship between the two different grading methods: traditional total points and assessment-based grades received both before and after retakes, and MAP standardized test score values. The results of the correlation analysis outlined in Table 7 showed that there was a statistically significant moderate positive relationship between grades received through different grading methods and standardized MAP test scores.

Table 7

Pearson Correlation Between Students' Grades Derived From Different Grading Method and MAP Scores

		Assessment-Based	Assessment-Based Retakes	MAP Scores
Total Points	Pearson's *r*	0.985	0.971	0.454
	p-value	< 0.001	< 0.001	< 0.001
Assessment-Based	Pearson's *r*		0.986	0.400
	p-value		< 0.001	.002
Assessment-Based Retakes	Pearson's *r*			0.381
	p-value			.004

Total-points grades were positively related to standardized test scores ($r = 0.454$, $p < 0.001$). Similarly, positive but not as strong of a relationship was found between assessment-based scores with standardized MAP test score ($r = 0.400$, $p < 0.001$ before retakes and $r = 0.381$, $p < 0.001$ after retakes). These results provided a positive answer to the third research question demonstrating a positive relationship between all grading methods and standardized tests scores.

Grading method comparison to test scores. Comparison of correlation coefficients for grades derived from different grading methods and MAP test scores was performed by using Hittner et al.'s (2003) test for the two overlapping correlations, as the correlations were calculated for the same variable, based on dependent groups (as the correlation was measured for

the same group of students). The results of the test, as outlined in Table 7, showed that the Pearson correlation coefficient obtained for total points grading method ($r = 0.484$) was statistically significantly higher than those received for assessment-based method (Fisher's $z = 2.51$, $p = 0.006$) and assessment-based with retakes (Fisher's $z = 2.43$, $p = 0.008$). These results provided a negative answer to the fourth research question (see Table 8) showing that students' scores calculated using a traditional total-points grading method were better related to standardized math test scores than those derived from both assessment-based methods, both before and after retakes.

Table 8

Hypotheses for Research Question 4

Research question 4	Directional hypothesis	Null hypothesis
Which grading method (traditional total-points grading and assessment-based grading) has resultant grades that more closely align to students' standardized test scores?	There is a stronger positive relationship between the grades derived from assessment-based grading practices, standardized test score values for eighth-grade math students than between the grades derived from traditional total-points grading, and standardized test score values for eighth-grade math students.	There is no difference in the relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students and between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students.

Summary

This chapter outlined the research results and provided answers to the central research question of the study: Which grading method determines classroom grades for Grade 8 math students that most accurately align, to corresponding standardized math MAP test score values for Grade 8 math students? The grades for Grade 8 math students calculated using the traditional total-points grading method, the assessment-based grades considering grades with and without retakes were compared to corresponding standardized math MAP test score values. The results

of mean values comparison showed that all three grading methods had lower grades than the corresponding standardized math MAP test scores while traditional total points showed that the smallest difference compared to assessment-based grading methods. Mean grades derived from the assessment-based grading method including retakes did not differ from traditional total-points grades. There was a positive relationship between different grading method and standardized MAP math test scores. The traditional total-points grading method grades showed a stronger relationship to standardized MAP math test scores compared with scores based on assessment-based grading methods when analyzed both including and excluding test retake scores.

Chapter 5: Discussion and Conclusion

Introduction

The purpose of this research was to discover differences between grades derived from total-points and assessment-based methods of grading, and to compare each set of grades to standardizes MAP test scores for the same group of Grade 8 math students for one 2018 semester term. To that end, this research uncovered a great deal of evidence that supports the effectiveness of the total-point grading method. Chapter 5 will provide a summary and discussion of the results and conclusions from this research study. These findings will be linked back to the literature reviewed in Chapter 2, and recommendations will be made for future research.

This researcher conducted an empirical review of the current research on grading practices and identified a deficit in research on the accuracies of assessment-based grading in comparison to traditional total-points grading. Less research was available on the assessment-based grading, as it was a newer grading practice, which is why this study presented an opportunity to add to the growing research repertoire for assessment-based grading by comparing it to a total-points grading approach. This has been achieved in so far as the findings from this research study now contribute to an area where there was limited research available. While still in the minority grading practice compared to a total-points method, the movement towards standards-based grading practices including assessment-based grading has increased in popularity since 2013 (Townsley, 2018). There is currently a gap in research that compares assessment-based grades to standardized test scores; more research is needed to inform the recommendation for the promotion or discontinuation of assessment-based grading, a derivative of standards-based grading, as a classroom best-practice. This study used a quantitative

correlational approach to identify if there was a relationship between the eighth-grade student grades derived from both assessment-based and total-points grading practices in comparison to standardized MAP math test scores.

This study was guided by the central research question: Which grading method determines classroom grades for Grade 8 math students that most accurately align, with the strongest correlation, to corresponding standardized math MAP test score values for Grade 8 math students? Under this central research question, the specific research questions addressed in this study included:

1. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students?
2. What was the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students?
3. What was the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students?
4. Which grading method (traditional total-points grading and assessment-based grading) has resultant scores that more closely aligned to students' standardized test scores?

This chapter will review the summary of results and discuss those results in detail. Secondarily, it will discuss the results in relation to the literature. This chapter will review the limitations of the study and assess the implications of these results for future practice, policy, and theory. Finally, it will provide recommendations for future research.

Summary of Results

With regard to the first research question, comparing the difference between grades calculated from different grading methods, the results of the analysis showed that traditional total-points derived grades were significantly higher than those assigned through the assessment-based grading method (mean difference of 1.3 points was statistically significant as proved by t (55) = 7.31, $p < 0.001$). The assessment-based scores received after retakes did not show a statistically significant difference from total-points derived grades. This means that the inclusion of higher retake summative test scores raised students' assessment-based percentage grades to be closer to those from the total-points grading approach. The total-points grades and assessment-based grades with retake scores did not demonstrate a difference of statistical significance, which means that they were similar for most students.

With regard to the second research question, comparing the difference in grades compared to test scores, the results showed that all grades received from both traditional total scores and both assessment-based methods (before and after retake) differed significantly compared to MAP standardized test score for Grade 8 math students. The grades derived from both grading methods did not closely correspond to student MAP test scores. Table 6 in Appendix A shows grade summaries for each student included in the study population. Many students scored higher on the MAP math test, converted into a percentage value, than either classroom grade calculated from either grading method, including the consideration of higher summative retake scores. Moreover, it can be concluded that the difference between total-points calculated grades and standardized test scores was significantly lower than the difference between assessment-based grades and standardized test scores.

Regarding the third research question, relating to the relationship between tests and test scores, the results found a statistically significant moderately positive relationship between grades received through different grading methods and standardized MAP test scores. Specifically, total-points grades were positively related to standardized test scores ($r = 0.454$, $p < 0.001$). After condensing a century of research, Brookhart et al., (2016) found that teacher-assigned grades typically correlate about .5 with standardized measures of student achievement. The total-points grade to MAP test score correlation value of $r = 0.454$ from this researcher's study is near the typical value found by Brookhart et al., (2016).

With regard to the fourth research question, the Pearson correlation coefficient obtained for total-points grading method ($r = 0.484$) was statistically significantly higher than those received for assessment-based method (Fisher's $z = 2.51$, $p = 0.006$) and assessment-based with retakes (Fisher's $z = 2.43$, $p = 0.008$). These results provided a negative answer to the fourth research question showing that students' scores calculated using a traditional total-points grading method was better related to standardized MAP math test scores than those derived from both assessment-based methods (both before and after retakes). Of the classroom grades calculated from student scores in this study, the total-points grades had the strongest relationship of statistical significance with standardized test scores. On average, the mean of the total-points grades was higher than the mean of the assessment-based counterparts, both with and without consideration of retake scores. The Grade 8 math students' MAP scores were higher, on average, than the classroom grades. Hence, the total-points' grades were proven not to be inflated by an overemphasis on formative points or inclusion nonacademic factors. Based on these results, total-points derived grades were a better representation of cumulative student learning in comparison to standardized test scores.

Discussion of Results

In terms of the difference between grades derived from different grading methods, the students' assessment-based semester classroom grades were measured twice: before and after retakes. This was important to isolate the impact of retake scores from the formative and summative categorical weighting differences between grading methods. The total-points grades did not include any test retake scores and represented only the students' initial attempts in classroom assessments. Every hypothesis, including assessment-based grades, was checked twice: first for scores received before and then after inclusion of summative test retake scores. This was in an effort to align with Bloom's (1968) theory of mastery learning, which serves as the conceptual framework for this study, and includes the opportunity to re-take tests and re-demonstrate knowledge.

Research question one stated: What is the difference, if any, between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading for one semester for Grade 8 math students? The directional hypothesis was that there was a statistically significant difference between mean values of the semester classroom grades derived from two different grading methods, traditional total-points grading and assessment-based grading, for Grade 8 math students which was is supported by this study. However, the difference favored traditional total-points rather than assessment-based grading. The results of the test showed that traditional total-points were significantly higher than those received through the assessment-based grading method (mean difference of 1.3 points was statistically significant as proved by $t(55) = 7.31$, $p < 0.001$). The greater impact from formative homework scores on the final class grade, due to point-assignments, in the total-points grading approach lessened the effect of summative tests on the total-points classroom grade. This resulted in a higher mean

total-points grade for students when compared to the assessment-based grading method that weighted summative test scores as 90% of the class grade. The inclusion of higher summative test retake scores in the weighted categories for assessment-based grading raised those grades closer to, but not equal to, the average values derived using the total-points grading approach. In the assessment-based grading approach, the inclusion of higher summative classroom test retake scores resulted in grades more comparable to the total-points grades. Therefore, the assessment-based scores received after retakes did not show a statistically significant difference from total-points derived grades.

With regard to the difference in grades compared to test scores, the second research question pertained to what the difference, if any, was between the academic grades derived from the two different grading methods: traditional total-points and assessment-based grading compared to MAP standardized test score for Grade 8 math students. The directional hypothesis stated that there was a statistically significant difference between mean values of the semester classroom grades derived from traditional total-points grading and MAP standardized test score for Grade 8 math students and a statistically significant difference between mean values of the semester classroom grades derived from assessment-based grading and MAP standardized test score for Grade 8 math students. The results showed (Table 2) that all grades calculated using both the traditional total-points method and assessment-based method, before and after summative test retake scores, differed significantly compared to MAP standardized test score for Grade 8 math students. The greatest difference of -10.4 was between assessment-based grades without a retake and standardized tests score (t (55) = -10.14, $p < 0.001$), while the difference between assessment-based grades with retakes compared with test scores was lower -9.3, but

also highly statistically significant (t (55) = -9.36, $p < 0.001$). The least difference, but also highly significant, was between traditional total-points grades and standardized test scores (M = -9.0, t (55) = -9.56, $p < 0.001$). Taking into account that there was a statistically significant difference between total-points and assessment-based grades before retakes, it can be concluded that the difference between total-points calculated grades and standardized test scores was significantly lower than the difference between assessment-based scores and standardized test scores. These results answer the second research question showing significant lower grades for all grading methods compared to standardized MAP test scores with the smallest difference obtained from total-points grades.

With regard to the relationship between grades and test scores, research question three sought to determine, what is the relationship, if any, between the grades derived from two different grading methods: traditional total-points grading and assessment-based grading, and MAP standardized test score values for Grade 8 math students. The directional hypothesis was that there is a positive relationship between the grades derived from assessment-based grading practices and standardized test score values for eighth-grade math students and there is a positive relationship between the grades derived from traditional total-points grading and standardized test score values for eighth-grade math students. The results of the correlation analysis showed that there was a statistically significant moderately positive relationship between grades received through different grading methods and standardized MAP test scores. This was important as it provided evidence that both methods of classroom grading produced math grades for the study population were at least moderately related to corresponding student MAP test scores. Based on this generalized relationship, a more specific analysis was conducted to quantify the strength of each relationship between grades and standardized test scores. Total-points grades were

positively related to standardized test scores ($r = 0.454$, $p < 0.001$). Similarly, positive but not as strong a relationship was found between assessment-based scores with standardized MAP test score ($r = 0.400$, $p < 0.001$ before retakes and $r = 0.381$, $p < 0.001$ after retakes). These results provided a positive answer to the third research question demonstrating a positive relationship between all grading methods and standardized tests scores.

Regarding the grading method comparison to test scores, research question four sought to determine which grading method (traditional total-points grading and assessment-based grading) has resultant grades that more closely align to students' standardized test scores. The directional hypothesis posited that there is a stronger positive relationship between the grades derived from assessment-based grading practices, standardized test score values for eighth-grade math students than between the grades derived from traditional total-points grading, and standardized test score values for eighth-grade math students. The results did not substantiate this and in fact showed that the Pearson correlation coefficient obtained for total-points grading method ($r = 0.484$) was statistically significantly higher than those received for assessment-based method (Fisher's $z = 2.51$, $p = 0.006$) and assessment-based with retakes (Fisher's $z = 2.43$, $p = 0.008$). These results provided a negative answer to the fourth research question showing that students' scores calculated using a traditional total-points grading method was better related to standardized math test scores than those derived from the weighted assessment-based method, both before and after retakes.

The findings for the first research question would be a higher difference for total-points compared to assessment-based practices but a difference for both was to be expected. The outcome of the results pertaining to the second research question with the differences in both total-points and assessment-based practices compared to standardized tests showed a lower

difference for the total-points grades. While the hypothesis was correct, the research and previous literature reviewed implied that assessment-based would be more accurate compared to total-points particularly specifically as it related to the concept of learning for mastery, which turned out not to be the case. The third hypothesis revealed a positive relationship between both methods of grading and standardized tests results. Most unexpected was the results of the fourth research question findings, which stipulated that a traditional total-points practice was more accurate, compared to an assessment-based grading practice in terms of its closeness to standardized MAP math test scores.

One explanation for the findings in the fourth research question could simply be that minimal research has been conducted pertaining to assessment-based grading methods; therefore, this is one of the first studies to explore this particular topic as a comparison revealing the potential downfalls and inaccuracies of the assessment-based grading practice. This researcher predicts that the mean of the assessment-based grades with retakes would have been higher if all students would have completed summative test retakes for a higher grade. Only about half of the participating students did at least one retake, and fewer than a third of the student population completed retakes on more than one of four summative assessments. The inclusion of summative test retake scores in the assessment-based grading method decreased the difference in comparison to total-points grades. It should be noted, however, that because test retakes are optional and not mandatory, only 30 of the 56 students chose to take a test retake that resulted in a higher score. Conversely, 26 of the 56 students did not take a single test retake, so their two weighted assessment-based grading percentages were identical. This is documented in the grade summaries chart in Appendix A. Consequently, it cannot be stated with certainty that all assessment-based grades explicitly included mastery-driven instructional strategies from the

participating teacher for this study. This researcher is unable to predict the potential correlational impact if all participating students were to have taken test retakes, but recommends this for consideration in future research.

Another possible explanation for the results for research question four may stem from some commonalities that are found in all grading systems. Assessment-based grading is an offshoot of standards-based grading, which means that it is not entirely unrelated to a total-points method of grading as there is a natural weighting of test grades in a total-points based approach through the emphasis on tests through points determination in comparison to assignments. Both grading systems use different mechanisms to emphasize summative performances, which most educators agree should be the predominant component of classroom grades (Wormeli, 2018). Formative and pre-assessments diagnostically guide instruction, while the summative assessments serve as a summative judgement of learning at the end of the learning cycle (Wormeli, 2018). There are different grading strategies because educators differ in their expectations of when the learning cycle ends as it serves grade determinations. The traditional total-points grading approach has the longest tradition within the education system based on a survey from 1998, O'Connor (2009) that found that 91% of schools at that time still used this traditional grading system. This would explain why the review of current literature found positive findings in support of both grading methods and yet slightly more favorable findings for the total-points practice, as total-points is still the predominant grading method with more readily available research.

Discussion of Results in Relation to the Literature

Some researchers supported the usage of a traditional total-points grading strategy (Akins, 2016; Comes, 2015; Denson, 2013, Fink, 2015; Lee, 2013) and their findings were all

corroborated by this research study. Denson (2013) conducted a correlational, non-experimental research design to examine the relationship between teacher-assigned grades and students' test scores. Denson (2013) confirmed the effectiveness of using a total-points grading strategy, as it accurately reflected students' classroom proficiencies comparable to students' performances on the FCAT. Fink (2015) designed an action research study to compare a standards-based grading practice to a traditional, total-points approach in respect to student achievement. On average, the students in the traditionally graded sections had a higher class-average than those classes with standards-based grading at 80% and 76%, respectively (Fink, 2015). In terms of academic achievement as represented by a semester grade, students in the traditionally graded total-points sections had, on average, a greater academic achievement. Lee (2013) also evaluated the accuracy of traditional classroom grades of students in correlation to standardized test scores. The researcher applied a Spearman rho test to the student data and identified a strong positive correlation, with a correlation coefficient of .82, between teacher-assigned classroom grades and standardized test scores (Lee, 2013). This provides evidence of the accuracy of a traditional total-points classroom grading practice and corroborates the findings of this research study as it relates to a traditional total-points approach being more effective than assessment-based grading in its closeness to standardized test results. More specifically, having a positive relationship to standardized test results, and having a lower difference between the total-points grades determined by instructors versus standardized test scores.

Pekel (2013) conducted a similar study that compared total-points grades and assessment-based grades to standardized test scores for Grade 8 math students. In 2007–2008 Pekel (2013) found the correlation coefficient between traditional total-points student grades and state standardized test scores was nearly negligible at 0.00 to 0.20 (Pekel, 2013). After implementing

an assessment-based grading practice in 2008–2009, a stronger correlation of 0.40–0.60 was found between student assessment-based grades and standardized test scores. A delimitation of this study was that there were two different student populations. The two participating teachers remained the same, but each had different students enrolled for the subsequent academic year. This makes it difficult to directly attribute the greater correlational results to the change in instruction and grading from total-points to assessment based, as there was another factor of different students. It is acknowledged that Pekel's (2013) study may be interpreted as weak, but serves as a reference point was there is a limited body of research comparing total-points grading and assessment-based grading. This study also found a moderately positive relationship between assessment-based grades and standardized MAP math test scores for this student population. Consequently, while this study found that total-points grades had a more statistically significant relationship, it does not directly contradict the results from Pekel. The findings of this study are similar, insofar as there exists a small difference, corroborating the work by Pekel.

While this study did not conduct longitudinal evaluations, the longevity of use of traditional total-points grading practices makes it a comfortable choice for many educators (O'Connor, 2009). Akins (2016) demonstrated the preference for a traditional total-points grading system by high school administrators. Akins (2016) conducted a descriptive, non-experimental quantitative survey of 247 lead high school principals from across the state of Missouri to evaluate themes in perception about the traditional total-points grading practice. Participants were given an attitudinal survey to assess the perceptions of the traditional grading practice from secondary principals. Of those provided with the survey across the state of Missouri, 33% agree with traditional grading practice of determining a term grade through averaging assignments (Akins, 2016). The results from Akins (2016) showed the preference

from high school principals for the use of a traditional total-points grading practice over other available grading procedures. The results from this researcher's study, including an evaluation of both total-points and assessment-based grading, add to the body of research that reinforces the use of total-points grading, the grading system that is widely used and preferred by secondary administrators according to Akins (2016).

For standards-based grading, Pollio and Hochbein (2015) used analysis to find the correlation coefficient for each group between the total-points grades and standards-based grades of secondary math students and KCCT standardized math test scores. The group with the standards-based math grades was found to have the highest correlation to KCCT math test scores. Total-points assigned grades were found to only have a weak correlation to KCCT test scores (Pollio & Hochbein, 2015). Assessment-based grading is a derivative of standards-based grading as an outcome-based grading approach. So, while not identical in methodology to Pollio and Hochbein (2015), it could be expected to find similar results between assessment-based and standards-based grading as both employ mastery-driven instructional practices. The results from this researcher's study found a moderate correlational relationship between total-points grades and standardized MAP math test scores over assessment-based grades. This researcher's analysis contradicts the findings from Pollio and Hochbein (2015).

In a qualitative study, Hardegree (2012) provided an analytic comparison of the standards-based grade cards, or SBRC, of 550 fifth-grade students in Georgia as opposed to individual standardized test scores from a criterion-referenced competency tests in math and reading. Both evaluation measures placed students into categories if they did not meet, met, or exceeded the standards. The results of the study showed that the students scored higher on the standardized math test than on the SBRC; the SBRC underestimated the test scores for most

students (Hardegree, 2012). This is substantiated by the findings of this research study, which indicated that performance on standardized tests is higher compared to the results from either total points or assessment-based grading; however, that there was less difference between total-points and standardized scores, comparatively.

Beatty (2013) presented data that standards-based grading is also successful for students at the postsecondary level. Beatty (2013) explained how standards-based grading was a more accurate representation of student learning, promoted deeper understanding and learning even after initial failures. The research represents a range in student ages, demonstrating the overall effectiveness of standards-based grading and its applicability in all levels of education. As such, the findings of total-points as a more accurately aligned form of grading as it relates to standardized tests scores aligns with the work of Beatty insofar as it could theoretically be applicable beyond the eighth-grade population used for this researcher's study.

The results comparing the difference between grades derived from different grading methods indicated that the assessment-based scores received after retakes showed no statistically significant difference when compared to the total-points' derived grades. Bloom (1968) asserted that, given the proper circumstances that extended the instructional time as well as the opportunities for instructional games, over 90% of students would be able to learn course content or learning objectives at a mastery level. The key factors to increasing this instructional time specifically for undemonstrated learning objectives included providing students with the opportunity to raise their grades from a retake, or a retest. This research study evaluated the comparative analysis of the accuracies for traditional total-points versus assessment-based grading practices and the findings support that total-points grading is significantly higher in terms of results and when combined with the fact that summative test retake scores, included as a

key component of a learning-for-mastery approach, had no statistically significant difference for the assessment-based grading method. These findings stand in direct opposition to Bloom's notion that mastery, as represented by a classroom percentage grade, could only best be accomplished through lengthened instructional windows of time, including opportunities for re-demonstration of learning, as with test retakes, and additional instruction. This researcher's analysis found total-points grades that did not explicitly include mastery-driven instructional strategies were found to more closely correlation to students' standardized MAP math test scores that served as a cumulative summative benchmark.

If grading practices were standardized through the implementation of a single grading policy, the assigned grades would become more meaningful and accurate for comparisons (O'Connor, 2012,) which can only be achieved effectively is teachers and instructors know which method is best. The results from this researcher's correlational study suggest it may be the total-points grading system. There are pitfalls of all grading practices to be mindful of. It is not uncommon for academic grades to be influenced by student behaviors, distorted evaluations, poorly organized evidence, and or inappropriate calculations including zeros that may skew an accurate representation of student learning (O'Connor, 2012). Grades are not essential for learning to occur; however, student grades play a central role in the education system and can provide feedback towards curriculum decisions and of instructional effectiveness (Brookhart et al., 2016). Academic grades can be used to make recommendations for students for certain degree programs or career paths (Guskey, 2011). Students can use grades for self-evaluation and goal setting (Wormeli, 2018). As a consequence of their prominent role in the education system, it is imperative that the most effective grading practice is selected for implementation. This research, revealing how closely aligned a total-points grading system was with relation to

standardized test results, now provides data which may support which grading practice is indeed the best choice for use by classroom teachers.

A traditional total-points grading practice, when viewed through the lens of a mastery-learning approach, was susceptible to being less effective overall because of the limited learning and testing window offered to students. Based on the findings by this researcher's study, this proved not to be the case, as the total-points grading practice was more effective through statistical analysis. More specifically, the results pertaining to research question two answers that second research question, showing significantly lower grades for all grading methods compared to standardized MAP test scores with the smallest difference obtained from total-points grades. The results of the test showed that traditional total points grades were significantly higher than those received through the assessment-based grading method (mean difference of 1.3 points was statistically significant as proved by t (55) = 7.31, $p < 0.001$), while the assessment-based grades calculated after retakes did not show a statistically significant difference from total-points derived grades.

Saefurrohman and Balinas (2016) indicated the retake component of the assessment-based grading approach, or opportunity to increase learning and mastery then demonstrate that newly gained knowledge, was conducive to a mastery learning approach. This was because it elongated the instructional period, provided individualized feedback after an evaluation with specific reteaching and an opportunity for students to further demonstrate their content knowledge through a summative assessment. The findings from this research study did not corroborate this, but do not discount the potential benefits of additional personalized instruction and an extended learning window as promoted through mastery-learning strategies. The longstanding body of research in support of mastery-learning approaches speaks to the

imperative influence of learning for mastery as an important factor for overall achievement. Because many teachers include diagnostic formative evaluations with responsive instructional changes, it is likely to assume that mastery techniques could be unintentionally incorporated into the traditional total point's instructional methods as well. It is not possible to determine the extent of any inclusion of mastery-driven instructional methods through a correlational analysis of archival data in this study; however, this factor could be addressed through future research with an experimental methodology.

Limitations

There are potential issues, as with any standardized test, of students exerting minimal or no effort when taking the MAP math test. This would result in an inaccurate assessment of that student's content knowledge, which would consequently prove to be an erroneous benchmark with which to compare classroom grades. The same limitations would apply to any standardized test. This researcher found, on average, MAP test scores for this student population were higher than their corresponding classroom grades, which suggests that this limitation had a negligible impact in this study.

Additional limitations refer specifically to the sample size. The number of units of analysis used in this sample size was limited by the number of students enrolled with the participating teacher. Previous grading research studies have undertaken hundreds of participants in a single study, whereas this study had fewer than 100 participants. The sample size for this study was restricted because the researcher only had access to student data from one participating teacher. On that note, a larger sample size would be necessary to ensure a representative distribution of the population or to have the results generalized or transferred to a larger number of schools and districts.

Having limited prior research to serve as the foundation for understanding the context and implications of the problem under investigation limited comparisons for the results of this study. However, having limited research served as an important opportunity to identify new gaps in the literature, nonetheless it is considered a limitation. The time available to investigate this research problem and to measure changes over time was constrained by the due date and as such while the results are valid, they are limited to one specific point in time and one specific set of grades therein. Due to these acknowledged limitations, this research study may best serve to add to the current limited body of literature on the effectiveness of assessment-based grading, and to provide recommendations for further research.

Implications of the Results for Practice, Policy, and Theory

The findings from this study may be important for consideration when making grading decisions regarding practice, policy, and theory. The results in answer to research question one, which indicated that a total-points method is more effective in terms of its significance and a lack of any statistically significant difference for assessment-based grades after students completed the retake, would suggest that mastery can be achieved with the total-points method of grading. For this study, it was assumed that mastery-learning strategies were not purposefully or explicitly incorporated into the total-points grading practice. And yet, despite its not being purposefully incorporated it stands to reason that mastery was achieved with the same level of success as those implementing the assessments-based grading practice. These results challenge the importance of offering summative test retakes for students, a practice that inevitably increases the workload on the classroom teacher. Furthermore, it warrants further discussion on the cost benefit analysis of summative test retake opportunities for students.

The results from this study should be considered in the context of the theory of mastery learning. While most stakeholders agree that the goal of instruction is student mastery of the learning objectives, there was still ambiguity as to the measure of the impact that a grading practice has on student mastery. The volume of research that supported Bloom's (1968, 1974, 1976, 1982, 1984) theory of mastery learning confirmed the effectiveness of mastery-driven strategies. The results from this study suggested principles of mastery learning could still be intentionally incorporated into instructional practices within a total-points grading system. This researcher does not dispute the soundness of Bloom's (1968) theory of mastery learning. Instead, this researcher suggests that mastery-learning tactics may be so intrinsically incorporated into instructional strategies that mastery can be achieved from multiple grading practices. Consequently, this study provides evidence that mastery may also be achieved using a total-points grading practice.

Students face a great deal of inconsistency in terms of the grades they receive. A single student could receive two different letter grades from two different teachers for the same performance (Wormeli, 2009). Grade values are important especially in comparison to the testing benchmark standard. According to Brookhart (2011), a student's grade should represent mastery of corresponding content learning objectives. Having the most accurate representation of grades is instrumental in setting students up for success. School officials rely upon those grades to rank students, to differentiate which students should be awarded college acceptance, and to document a satisfactory level of learning in classes (Brookhart, et al., 2016). Grades are also used to award course credits at the collegiate level and to distinguish exemplary students (Schneider & Hutt, 2013). As a result, academic grades have the potential to influence a

student's future and provide him or her with opportunities or possible limitations with post-secondary opportunities.

If students are given grades that reflect inaccurately upon their abilities or achievements, it could manifest in a student missing out on opportunities because those grades indicated subpar performance when this was in fact not true. Similarly, students whose grades indicated excellent performance when this was in fact not true might receive credit for something they have not truly mastered or receive opportunities for which they are ill-prepared. Grades need to be calculated with the best possible approach because of this significance and potential consequences (Muñoz & Guskey, 2015). Instructors have a responsibility to accurately assign grades in the most meaningful way and in order to help them do that they need to know which of these two prominent grading systems is most closely aligned with standardized test results so that students can receive a clear reflection upon their abilities as it aligns directly with state standards. The results of this study indicated that the traditional total-points grading approach is the most effective and closest in relation to standardized test scores. As a result, instructors and educators should consider implementing or keeping a total-points grading method in order to provide students with the most accurate grades as a representation of learning.

On a larger scale, grades would be more comparable across and between school districts if there was more consistency in how classroom grades were calculated. Implementation of standardization measures should be administered not just across individual classrooms, but also across entire school districts by principals and administrators alike. Having standardized implementation would do away with any confusion or ambiguity pertaining to grades and would alleviate the issue of children receiving grades that are different from different teachers for the same performance. This would also make it easier for college admissions committees and other

similar organizations to make more accurate decisions about grading influenced decisions such as post-secondary opportunities, college acceptance, class rank, and scholarships. It is not enough to have teachers simply using the same grading method in theory, the tenets of the grading practice must be upheld with fidelity and consistency to achieve a standardized practice. This feat will require standardizations of grading practices that cannot be accomplished without providing specific expectations and purposeful structure and support for teachers. Such trainings may require district-wide professional development opportunities to help foster an understanding in all participating teachers on how to implement the grading program with fidelity.

Many school districts, including the district that provided the setting for this study, are now requiring teachers to use assessment-based grading for learning practices as a school-wide grading policy (XXXXX & XXXXXXXX, 2017). This host district did provide professional development sessions to teachers, specifically regarding minimum weighted category values and mandated summative retake opportunities, when it changed from total-points grading to requiring teachers to use assessment-based grading practices to ensure compliance and consistency school-wide. Assessment-based grading grew in popularity starting in the early 2000s (Kalnin, 2014). If a school district is going to require a teacher to use a mandated grading policy, it is most responsible to promote a research-supported grading practice.

This study has identified a relationship between assessment-based grades and standardized test scores. Having found that the total-points method is more effective than the assessment-based practice, many school districts, including the district that provided the setting for this study, should reconsider implementation of school-wide assessment-based grading policies especially in light of the findings. Given the results of this study, administrators and stakeholders in the field of education would do well to assess their local grading practices and

should not discount the effectiveness of a total-points grading practice. While the task of instruction is never effortless, there are facets of assessment-based grading instructional approaches, including additional personalized intervention reteaching and reassessment, that can require substantially more work for the classroom teacher. The results from this study in support of a total-points approach that does not necessarily require the reteaching and retesting loop yet is still a more accurate representation of cumulative student learning in comparison to standardized test scores than the assessment-based approach.

In the literature, there was division among principals who reported a support for grading practices included in the traditional total-points approach in the study by Atkins (2016). Around 35% of principals reported that late work should be penalized through a point deduction and 40% instruct teachers to enter scores of zero for missing work (Akins, 2016). The discrepancy in terms of late work penalizations and zero points for missing work is something that was theoretically improved upon with the assessment-based method but could stand to be better implemented in the total-points' method in such a way that overall grades can be improved for those who submit late work. In a total-points approach formative scores can count for more in the overall grade calculation because they are not capped by a minimally weighted category as in the assessment-based approach. Consequently, in a total-points method a zero score for missing work has the potential to more heavily skew the overall grade. Likewise, point deduction penalizations for late work submissions may be more influential in a total-points grading practice. This too should be a consideration in future research.

In the study by Merki and Holtmeier (2015), participating teachers used a total-points grading approach; however, each had different point-assignments to different categories and assignments, resulting in inconsistent classroom grade determinations. This is a susceptibility

for all total-points grading practices, as there are no set recommendations for the proportion of point assignments for any category of constituting formative and summative scores. Each teacher applies his or her own approach to point-assignments, which consequently weights different assignments within the total-points system. In a total-points grading system, the points available for participation and homework are ~~is~~ usually enough for low-achieving students to earn a passing grade (Akins, 2016). In a total-points system, students who are responsible and have a good work ethic may earn an inflated, or at least passing, grade in a class without demonstrating high levels of learning of the learning objectives and standards (Akins, 2016). That said, were a comprehensive total-points approach to be widely implemented as the standard grading practice across any given institution or district, it is recommended that a protocol be determined, communicated and implemented with teachers to standardized point-assignment expectations to ensure consistency between teachers.

The results from this study endorse the practice of a total-points grading strategy in classroom grading. Individual districts may select a single grading policy for all staff members to use for classroom grading so that there is consistency in grade determinations. It may not be the grading practice that, but instead the teacher's instructional strategies that better facilitate mastery-learning for students. The participating teaching in this study has taught math for over 30 years. As an experienced educator, it is assumed that the participating teacher used formative assessments diagnostically in efforts towards facilitating student mastery. A school district may mandate mastery-driven methods for teachers through the requirement of an assessment-based or standard-based grading practice, which explicitly includes a reteaching and reassessment loop. This policy may best serve classroom teachers that are less effective at inherent inclusion of mastery practices during the formative cycle.

Recommendations for Further Research

In this study, the participating teacher did not explicitly use mastery-learning strategies in the total-points grading; however, assessment-based grading did incorporate the key components of Bloom's (1968, 1974, 1976, 1982, 1984) theory of mastery learning. There was an individualized reteaching and reassessment loop with the participating teacher in the assessment-based grading method that served explicitly to facilitate mastery using Bloom's (1968) recommendations. Tangentially, one of the most unexpected results was the fact that allowing students to retake tests for the assessment-based grading practice showed no statistically significant difference in grades when compared to the total-points system. It was theorized that achievement of mastery could only effectively be accomplished for the assessment-based grading method since that method purposefully incorporated the approach outlined by Bloom (1968) as necessary for mastery.

Brophy (2013) focused on the impact of student motivation, specifically resulting from retake opportunities, to increase mastery. Brophy (2013) found that student motivation was related to grades and the amount of content mastered. Awarding academic grades, specifically with an opportunity to improve them, supported a behaviorist's approach to teaching and provided feedback for students, which increased student motivation to work toward the desired outcome of mastery. These factors of the goal theory are conducive to the tenets of mastery learning. Additionally, Fink (2015) supported a total-points based approach; however, surveys and interviews from 25 participating students reported an increased student preference for assessment retake opportunities in the standards-based approach.

Given Fink's (2015) finding of student preference for retake opportunities and Marzano and Heflebower's (2011) recommendations for repeated assessment opportunities for students

over time, the potential value in providing additional assessment opportunities for students to better demonstrate knowledge gains should not be ignored. In light of the small difference in grade means from this researcher's study, it stands to reason that future research could expound upon this particular area and determine what purposeful measures towards mastery may or may not be necessary to incorporate into the total-points grading practice in order to heighten student proficiency. Based on the results of this study, future research that addressed this consideration may provide the distinction between mastery instructional methods and the effects of categorical weighting when determining classroom grades.

Given this study's smaller sample size, future researchers' studies would do well to recreate this study but simply using a much larger group of students. Additionally, this could relate to additional grades, not just eighth-grade, as well as additional standardized tests subjects, not just math. Future studies with a larger sample size of Grade 8 math or any other subject/ grade would produce results more easily generalizable or transferable to a much larger group of students, statewide or even nationwide. Future researchers' studies could also capitalize upon the time constraints of this particular research by conducting a longitudinal study to review long-term effects of both assessment-based grading methods compared to total-points grading methods. While the findings of this study substantiate the fact that total-points grading is more effective and should be implemented overall, it stands to reason that future research could conduct much longer comparisons to verify whether these results hold true long-term.

Conclusion

Overall, this researcher conducted an empirical review of the current research on grading practices and identified a deficit in research on the accuracies of assessment-based grading in comparison to traditional total-points grading. O'Connor (2012) promoted an increase in

districts using assessment-based grading has increased in popularity; however, there is currently a gap in research that compares assessment-based grades to standardized test scores. More research is needed to inform the recommendation for the promotion or discontinuation of assessment-based grading as a classroom best practice. This researcher's study used a quantitative approach to identify if there is a relationship between the eighth-grade student grades derived from both assessment-based and total-points grading practices in comparison to standardized test scores. The findings uncovered that there is a high correlation between total-points grades and standardized test results, higher than the findings relating to assessment-based grading results. As a result, it is recommended that institutions including that which provided the setting for this study, reconsider the implementation of assessment-based grading methods and return to the older method of total-points grading as it provides a more accurate representations of student grades.

References

Adrian, C. A. (2012). *Implementing standards-based grading: Elementary teachers' beliefs, practices and concerns* (Doctoral dissertation, Washington State University). Retreived from http://hdl.handle.net/2376/4090

Akins, J. (2016). *Secondary principals' perceptions of grading and grade reporting practices* (Doctoral dissertation, Southwest Baptist University). Retrieved from ERIC Institute of Education Sciences database.

Beatty, I. D. (2013). Standards-based grading in introductory university physics. *Journal of the Scholarship of Teaching and Learning, 13*(2), 1–22.

Beltzner, D. (2018). *What parents need to know about changes in mathematics education*. New York, NY: Scholastic Inc. Retrieved from https://www.scholastic.com/parents/school-success/school-life/grade-by-grade/what-parents-need-to-know-about-changes-mathematics-education.html

Block, J., Airasian, P., Bloom, B., & Carroll, J. (1971). *Mastery learning: Theory and practice.* New York, NY: Holt, Rinehart and Winston.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*. *1*(2). Retrieved from http://programs.honolulu.hawaii.edu/intranet/sites/programs.honolulu.hawaii.edu.intranet/files/upstf-student-success-bloom-1968.pdf

Bloom, B. S. (1974). Time and learning. *American Psychologist, 29*(9), 682–688. doi:10.1037/h0037632

Bloom, B. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.

Bloom, B. S. (1982). *All our children learning: A primer for parents, teachers, and other educators*. New York, NY: McGraw-Hill.

Bloom, B. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership, 41*(8), 4–17. doi:10.2307/1175554

Brandt, R. (2018). *An overview of outcome-based education. Curriculum Handbook*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD). Retrieved from www.ascd.org/publications/curriculum_handbook/413/chapters/An_Overview_of_Outcome-Based_Education.aspx

Brookhart, S. (2011). Starting the conversation about grading. *Educational Leadership, 69*(3), 10–14.

Brookhart, S., Guskey, T., Bowers, A., McMillan, J., Smith, J., Smith, L., … , Welsh, M. (2016). A century of grading research: Meaning and value in the most common educational manner. *Review of educational research, 86*(4), 803–848.

Brophy, J. E. (2013). *Motivating students to learn*. New York, NY: Routledge.

Carifio, J., & Carey, T. (2009). A critical examination of current minimum grading policy recommendations. *The High School Journal, 93*(1), 23–37.

College Board. (2018). *Admission decisions: What counts.* New York, NY: Self. Retrieved from https://professionals.collegeboard.org/guidance/applications/decisions

Comes, A. B. (2015). *Criterion-referenced grading vs. traditional grading: Using grades to predict summative test performance* (Doctoral dissertation, Concordia University Chicago).

Denson, B. C. (2013). *Examining teacher-assigned math grades as predictors for high-stakes testing* (Doctoral dissertation, Capella University). Retrieved from ProQuest database.

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE, 10*(4), e0121945. doi:10.1371/journal.pone.0121945

Educational Research Institute of America. (2009). *A longitudinal analysis of state mathematics score for Florida schools using Saxon Math*. (Report No. 365). Retrieved from http://www.curiousgeorge.com/~/media/sites/home/education/global/pdf/resources/mathematics/elementary/saxon-math/sxnmath_fl_effectivenessreport09-6-14-13.pdf

Fink, A. M. (2015). *A comparative analysis of a standard based grading system versus a traditional based grading system with respect to student academic achievement and motivation* (Doctoral dissertation). Retrieved from https://commons.cu-portland.edu/

Goodrich, K. (23 July 2012). *What Is Formative Assessment?* Portland, OR: Northwest Evaluation Association (NWEA). Retrieved from www.nwea.org/blog/2012/what-is-formative-assessment/

Great Schools Partnership (2014). *Criterion-referenced testing*. Portland, ME: Self. Retrieved from https://www.edglossary.org/criterion-referenced-test/

Grinberg, E. (2014, April). Ditching letter grades for a 'window' into the classroom. *CNN*. Retrieved from https://www.cnn.com/2014/04/07/living/report-card-changes-standards-based-grading-schools/index.html

Guskey, T. (2005, April). *Formative classroom assessment and Benjamin S. Bloom: Theory, research, and implications*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada. Retrieved from https://files.eric.ed.gov/fulltext/ED490412.pdf

Guskey, T. (2011). Effective grading practices. *Educational Leadership, 69*(3), 16–21.

Guskey, T. (2014). Class rank weighs down true learning. *Phi Delta Kappan*, *95*(6), 15–19.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008, December). *Standards-based reform in the United States: History, research, and future directions*[Scholarly project]. Center on Education Policy. Retrieved from https://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND_RP1384.pdf

Hanover Research. (2011*). Effective grading practices in the middle school and high school environments*. Washington, DC: Self. Retrieved https://njctl-media.s3.amazonaws.com/uploads/Effective grading practices in the middle school and high school environments.pdf

Hardegree, A. (2012). *Standards-based assessment and high stakes testing: Accuracy of standards-based grading* (Doctoral dissertation, Liberty University).

Heflebower, T., Hoegh, J. K., Warrick, P. B., Hoback, M., McInteer, M., Clemens, B., & Marzano, R. J. (2014). *A school leader's guide to standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, *130*(2), 149–68.

Houghton Mifflin Harcourt. (2018). *Saxon Math*. Boston, MA: Self. Retrieved from https://www.hmhco.com/programs/saxon-math

Houghton Mifflin Harcourt. (2019). *Middle grades Saxon Math placement test*. Boston, MA: Self. Retrieved from https://www.hmhco.com/~/media/sites/home/education/global/pdf/placement/mathematics/k-12/saxon-math-homeschool/sms_plt_middlegrades.pdf?la=en

Iamarino, D. L. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education, 17*(2), 1–12.

Kalnin, J. S. (2014). Proficiency-based grading: Can we practice what they preach? *AILACTE Journal, 11*(1), 19–36. Retrieved from http://files.eric.ed.gov.cupdx.idm.oclc.org/fulltext/EJ1052571.pdf

Knaack, S., Kreuz, A., & Zawlocki, E. (2012). *Using standards-based grading to address students' strengths and weaknesses* (Master's thesis, Saint Xavier University).

XXXXX, N., XXXXXXXX, D. (2017, August). MIDWESTERNjr/sr high school student/parent/guardian and activities handbook. Retrieved from https://docs.google.com/document/d/1wzgzYgLb6-bXXXXXXXXgoiP9MASBIB-6ZKbIcP-HrLw/edit

Lee, M. (2013). *Relationships between teacher-assigned grades and academic achievement as determined by AIMS testing* (Doctoral dissertation, Northern Arizona University).

Marzano, R. J. (2010). *Formative assessment and standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

Marzano, R., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership, 69*(3), 34–39.

Merki, K., & Holmeier, M. (2015). Comparability of semester and exit exam grades: Long-term effect of the implementation of state-wide exit exams. *School Effectiveness and School Improvement, 26*(1), 57–74.

Midwestern Department of Education (2017). *MIDWESTERNDistrict Profile 2017*. Midwestern, XX: Self. Retrieved from https://XXX.education.XX.gov/

Midwestern Department of Education (2018). *Midwestern's college and career ready standards for mathematics 2015*. Midwestern, XX: Self. Retrieved from https://2x9dwr1yq1he1dw6623gg411-wpengine.netdna-ssl.com/wp-content/uploads/2017/07/2015_XXXXXXXX_College_and_Career_Standards_for_Mathematics_Vertical.pdf

Mitee, T. L., & Obaitan, G. N. (2015). Effect of mastery learning on senior secondary school students' cognitive learning outcome in quantitative chemistry. *Journal of Education and Practice*, *6*(5), 34–38.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education: making clear linkages between standards, assessment, grading, and reporting that are concisely reported work for the betterment of ALL students. *Phi Delta Kappan, 96*(7), 64–68.

Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students* (6th ed.). Boston, MA: Allyn & Bacon.

NWEA (2018). *MAP growth*. Portland, OR: Northwest Evaluation Association (NWEA). Retrieved from https://www.nwea.org/map-growth/

O'Connor, K. (2009). Reforming grading practices in secondary schools. *Principal's Research Review*, *4*(1), 1–7.

O'Connor, K. (2012). How to grade for learning by using 15 fixes for broken grades. Retrieved from http://www.gfps.k12.mt.us/sites/default/files/Ken%20O'Connor%20GFPS%20Aug%2013,%202012.pdf

O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership, 69*(3), 40–44.

Pannucci, C., & Wilkins, E. (2010, August). Identifying and avoiding bias in research. *Plast Reconstr Surg., 126*(2), 619–625. doi: 10.1097/PRS.0b013e3181de24bc

Pekel, K. (2013). *A tale of two grades: An evaluation of grading for learning, A middle school grading reform* (Doctoral dissertation, University of Minnesota). Retrieved from https://conservancy.umn.edu/bitstream/handle/11299/36771/Pekel_umn_0130E_14502.pdf?sequence=1

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.

Rosales, R. (2013). *The effects of standards-based grading on student performance in algebra 2* (Doctoral dissertation). Retrieved from https://digitalcommons.wku.edu/diss/53/

Saefurrohman, S., & Balinas, E. S. (2016). English teachers classroom assessment practices. *International Journal of Evaluation and Research in Education, 5*(1), 82–92.

Set, A. (2018, February*). Study concludes MAP growth items align to common core state standards*. Portland, OR: Northwest Evaluation Association (NWEA). Retrieved from https://www.nwea.org/blog/2018/study-concludes-map-growth-items-align-common-core-state-standards/

Schinske, J., & Tanner, K. (2014). Teaching more by grading less. *CBE Life Sciences Education,13*(2), 159–166. doi:10.1187/cbe.CBE-14-03-0054

Schneider, J., & Hutt, E. (2013). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies, 46*(2), 201–224. doi:10.1080/00220272.2013.790480

Smeding, A., Darnon, C., Souchal, C., Toczek-Capelle, M., & Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS One, 8*(8), e71678. doi:10.1371/journal.pone.0071678

Spencer, K. (2017, August 11). A new kind of classroom: No grades, no failing, no hurry. *The New York Times*. Retrieved from https://www.nytimes.com/2017/08/11/nyregion/mastery-based-learning-no-grades.html

Stanford University (2000). *History of standards-based reform*. Palo Alto, CA: Self. Retrieved from https://web.stanford.edu/~hakuta/www/archives/syllabi/CalTex_SBR/historysbr.html

Starch, D. (1913). Reliability and distribution of grades. *Science, 38*, 630–636. doi:10.1126/science.38.983.630

Stefanowicz, M. (2016, January 4). *Navigating the college application process with your standards-based transcript*. New York, NY: Jumprope. Retrieved from https://www.jumpro.pe/single-post/2017/02/14/Navigating-the-College-Application-Process-with-Your-Standards-Based-Transcript

Summit Learning. (2015). *MAP data important terminology and concepts*. Redwood City, CA: Self. Retrieved from https://help.summitlearning.org/hc/en-us/articles/360008660253-MAP-Data-Important-Terminology-and-Concepts

Supovitz, J. (2018). *Is high-stakes testing working?* Philadelphia, PA: University of Pennsylvania Graduate School of Education. Retrieved from https://www.gse.upenn.edu/review/feature/supovitz

Timmons, D. (2017). *Challenging standards grading practices: A qualitative action research study* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/4087/

Tinkelman, D., Venuti, E., & Schain, L. (2013). Disparate methods of combining test and assignment scores into course grades. *Global Perspectives on Accounting Education, 10*, 61–80.

Townsley, M. (2018, February). Mastery-minded grading in secondary schools. *School Administrator*. Alexandria, VA: AASA, The School Superintendents Association.

Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. Alexandria, VA: ASCD.

Welsh, M. E., D'Agostino, J. V., & Kaniskan, B. (2013). Grading as a reform effort: Do standards-based grades converge with test scores? *Educational Measurement: Issues and Practice, 32*(2), 26–36.

Wormeli, R. (2009). *Grading and reporting for student learning* [Review of powerpoint Presentation by R. Wormeli]. Retrieved from https://bvsd.org/development/SBGDocuments/SBG%20Docs/Wormeli%20Grading-ReportingPresentation.pdf

Wormeli, R. (2018). *Fair isn't always equal: Assessment and grading in the differentiated classroom* (2nd ed.). Portland, OR: Stenhouse.

Appendix A: Grade Summaries

Table 9

Summary of Student Grades and MAP Score as Percentages

Student	TP	ABG	ABG (R)	RIT Conv.	Student	TP	ABG	ABG (R)	RIT Conv.
1	90.96	91.28	91.28	83.19	29	81.60	81.71	82.01	90.00
2	83.76	83.59	84.61	84.24	30	84.78	84.71	84.71	94.18
3	78.01	76.18	79.65	88.95	31	92.64	92.12	93.42	93.13
4	74.75	71.97	73.62	82.67	32	91.86	91.47	91.47	91.04
5	64.53	66.30	68.23	84.24	33	93.32	92.65	94.35	92.09
6	72.07	69.53	73.77	81.10	34	71.64	70.77	70.77	87.90
7	87.46	87.67	87.67	81.10	35	73.82	72.30	75.00	85.29
8	68.24	65.15	67.98	87.90	36	75.67	72.78	75.09	92.61
9	90.94	91.31	91.31	94.70	37	81.02	80.80	81.70	97.32
10	86.31	85.18	86.78	91.04	38	84.15	80.98	82.63	100.00
11	90.89	89.41	91.01	91.04	39	89.88	87.61	87.61	100.00
12	80.66	78.52	78.52	83.19	40	74.73	71.41	76.61	92.61
13	91.74	91.16	91.96	96.27	41	80.90	79.86	79.86	100.00
14	93.20	93.82	93.82	94.18	42	87.91	85.70	85.70	96.27
15	85.78	86.25	86.25	91.04	43	82.63	79.31	79.31	98.89
16	70.53	70.48	70.48	76.39	44	85.68	82.32	82.32	98.89
17	92.09	91.90	92.20	91.04	45	84.33	81.53	81.53	97.84
18	95.91	96.37	96.37	99.41	46	87.82	85.51	85.51	97.32
19	90.04	89.68	89.68	95.23	47	88.00	84.31	84.31	96.80
20	82.86	81.00	82.30	88.95	48	79.38	77.30	79.90	96.27
21	87.97	86.84	89.44	92.09	49	81.96	79.15	79.15	95.23
22	71.61	70.97	74.87	86.86	50	77.90	76.80	76.80	96.80
23	86.92	87.85	87.85	96.80	51	82.55	80.19	80.97	96.80
24	89.65	89.73	93.83	91.04	52	78.72	76.50	79.10	94.70
25	71.94	69.27	70.37	80.58	53	85.00	82.33	82.33	87.90
26	82.92	80.53	81.23	84.24	54	78.43	77.82	79.02	98.37
27	75.64	76.06	77.36	92.61	55	83.45	80.51	81.18	95.75
28	71.40	70.07	72.07	85.81	56	82.72	79.18	79.18	96.80

Appendix B: RIT Conversions

2015 MATHEMATICS Student Status Norms						
	Begin-Year		Mid-Year		End-Year	
Grade	Mean	SD	Mean	SD	Mean	SD
K	140.0	15.06	151.5	13.95	159.1	13.69
1	162.4	12.87	173.8	12.96	180.8	13.63
2	176.9	13.22	186.4	13.11	192.1	13.54
3	190.4	13.10	198.2	13.29	203.4	13.81
4	201.9	13.76	208.7	14.27	213.5	14.97
5	211.4	14.68	217.2	15.33	221.4	16.18
6	217.6	15.53	222.1	16.00	225.3	16.71
7	222.6	16.59	226.1	17.07	228.6	17.72
8	226.3	17.85	229.1	18.31	230.9	19.11
9	230.3	18.13	232.2	18.62	233.4	19.52
10	230.1	19.60	231.5	20.01	232.4	20.96
11	233.3	19.95	234.4	20.18	235.0	21.30

Figure 6. This figure outlines MAP RIT norm scores. The norm RIT value for Grade 8 students at the end of the spring semester is 230.9 with a standard deviation of 19.11. This figure is provided by NWEA and outlines MAP Math Norm Values (Summit Learning, 2015).

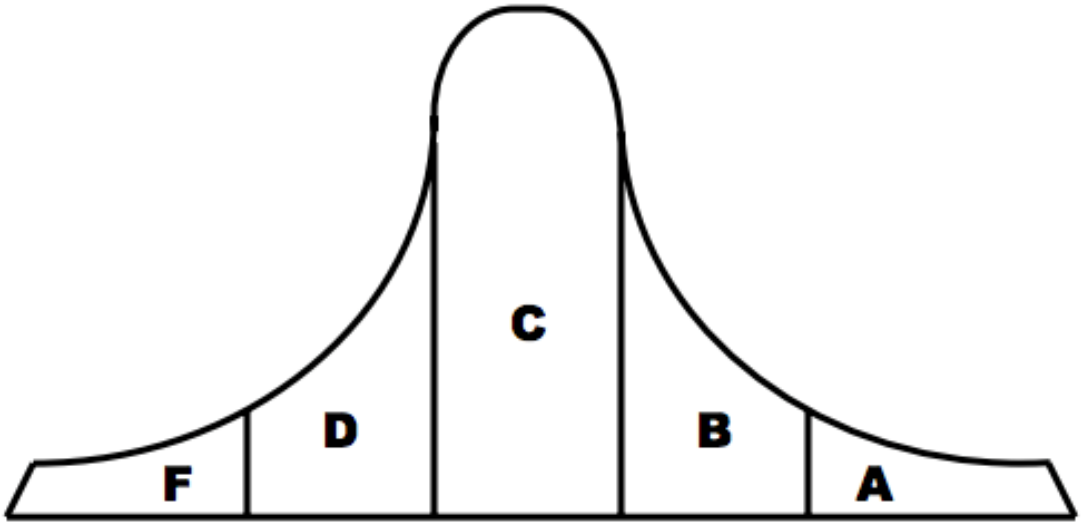

Figure 7. This figure illustrates the distribution of achievement in traditional classrooms (Bloom, 1976). This figure shows the distribution of student grades under a traditional bell curve. The letter grades are separated by a standard deviation, with grade assignments based on a traditional 60%–100%, A–F grading scale (Guskey, 2005).

Table 10

MAP RIT to Percentage Bell Curve Conversion Chart

RIT	%	RIT	%
269.12	100	230.9	80
267.209	99	228.989	79
265.298	98	227.078	78
263.387	97	225.167	77
261.476	96	223.256	76
259.565	95	221.345	75
257.654	94	219.434	74
255.743	93	217.523	73
253.832	92	215.612	72
251.921	91	213.701	71
250.01	90	211.79	70
248.099	89	209.879	69
246.188	88	207.968	68
244.277	87	206.057	67
242.366	86	204.146	66
240.455	85	202.235	65
238.544	84	200.324	64
236.633	83	198.413	63
234.722	82	196.502	62
232.811	81	194.591	61
		192.68	60

Appendix C: Statement of Original Work

The Concordia University Doctorate of Education Program is a collaborative community of scholar-practitioners, who seek to transform society by pursuing ethically-informed, rigorously- researched, inquiry-based projects that benefit professional, institutional, and local educational contexts. Each member of the community affirms throughout their program of study, adherence to the principles and standards outlined in the Concordia University Academic Integrity Policy. This policy states the following:

Statement of academic integrity.

> As a member of the Concordia University community, I will neither engage in fraudulent or unauthorized behaviors in the presentation and completion of my work, nor will I provide unauthorized assistance to others.

Explanations:

What does "fraudulent" mean?

> "Fraudulent" work is any material submitted for evaluation that is falsely or improperly presented as one's own. This includes, but is not limited to texts, graphics and other multi-media files appropriated from any source, including another individual, that are intentionally presented as all or part of a candidate's final work without full and complete documentation.

What is "unauthorized" assistance?

> "Unauthorized assistance" refers to any support candidates solicit in the completion of their work, that has not been either explicitly specified as appropriate by the instructor, or any assistance that is understood in the class context as inappropriate. This can include, but is not limited to:
>
> - Use of unauthorized notes or another's work during an online test
> - Use of unauthorized notes or personal assistance in an online exam setting
> - Inappropriate collaboration in preparation and/or completion of a project
> - Unauthorized solicitation of professional resources for the completion of the work.

Statement of Original Work (Continued)

I attest that:

1. I have read, understood, and complied with all aspects of the Concordia University–Portland Academic Integrity Policy during the development and writing of this dissertation.

2. Where information and/or materials from outside sources has been used in the production of this dissertation, all information and/or materials from outside sources has been properly referenced and all permissions required for use of the information and/or materials have been obtained, in accordance with research standards outlined in the *Publication Manual of The American Psychological Association*

Digital Signature

Emily A. Winter

Name (Typed)

June 7, 2019

Date