RELATIONSHIP BETWEEN GRADING PRACTICES AND TEACHER EFFICACY

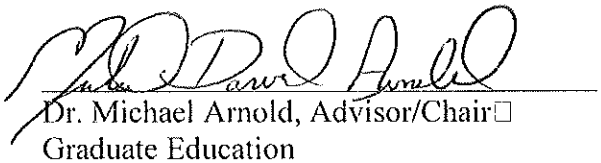
JASON WEAVER

2015

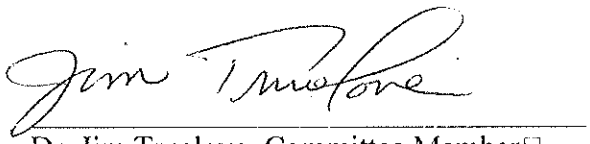
The undersigned, approved by the Department Chair of Graduate Studies in Education, have examined a dissertation entitled:

RELATIONSHIP BETWEEN GRADING PRACTICES AND TEACHER EFFICACY

Presented by Jason Weaver a candidate for the degree of Doctor of Education and hereby certify that in their opinion it is worthy of acceptance.



Dr. Michael Arnold, Advisor/Chair
Graduate Education



Dr. Jim Truelove, Committee Member
Graduate Education

Dr. Robert Perry, Committee Member
Graduate Education

RELATIONSHIP BETWEEN GRADING PRACTICES AND TEACHER EFFICACY

A Dissertation
Presented to
The Faculty of the Graduate Education Department
Southwest Baptist University

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Education

By

Jason Weaver, B.S., M.S., Ed.S.

Dr. Michael Arnold, Dissertation Advisor

November 2015

ACKNOWLEDGEMENTS

As my wife has observed in the past, there was never a doubt that I would pursue my doctorate at some point in my career, the only question was when. I have enjoyed the journey and learned a great deal about the content, leadership, and myself. I knew since my first day of kindergarten that I wanted to be a teacher when I grew up and with the same unwavering drive, I knew I would complete this process as a step to better myself as an educational leader. It was not just the process of studying and developing this dissertation but the people I formed relationships with during my time at Southwest Baptist University that have helped me grow as a leader.

I first and foremost want to thank my wife, Stephanie, for her support and encouragement. She was the financial guru who balanced the books and ensured we could afford the journey without incurring debt. She was the encourager who kept me going when I was deep into the research. She patiently listened as I thought aloud about my research, the process, my peers, and my own application of the skills I developed in my school. She was a mom to our boys when I was hidden away in the office writing. She made this journey possible and was a partner through it all. Stephanie, I love you.

My boys were also great encouragers. Watching Lucas work on his own "dic-er-tashun" just like daddy or taking extra breaks to read a book with Logan helped me maintain my drive to see the task through to the end. Though I hated the time spent on Saturdays missing our playtime to go to class or work on this paper, I am thankful for their unflinching love and support cheering for me each time I completed a phase of the journey. Daddy's almost finished and ready to get back to playing Super Friends.

Throughout the process, Dr. Mick Arnold, Dr. Jim Truelove, and Dr. Bob Perry have been great supports to me as well. I appreciated both the feedback on my research and the sense of humor these men had throughout our work together. Dr. Arnold especially sensed my desire to reach the end and was quick to return feedback and suggestions to keep me moving through the process. Dr. Truelove spent several hours helping walk through the statistics with me balancing Dr. Arnold's interest in the meaning with his own in the numbers so that I ended up with an understanding of what my survey showed. I am grateful for their time, feedback, and support. As promised, I'll make sure the time is worth more than just a burger and fries to you.

Lastly, I want to thank the three ladies who went from being colleagues to good friends during our rides back and forth to class each week. Jill White, Teresa Adams, and Amanda Boyer made the entire journey more than possible, they made it fun. From honest conversations in the car to late-night text strings, they have helped me keep my sanity and perspective not just in pursuing my doctorate but in becoming a better leader in my school. I still have so much more to learn from them and I am honored that they call me friend. Jill, T, and Amanda, you're in a three-way tie on my number one work-friends list.

DEDICATION

I've been blessed with many people in my life who have been here to cheer me on and support me through each step of my career. Of all the people, Jack Ross, my grandpa, has been one of the loudest.

His pride has not just been in my career, but in my life. He was proud to say his grandson, as a toddler, knew the name of every type of tractor we passed on our way to the little country store in Russ when he'd take me to get a treat. He was there with his camcorder to record every school performance and honor. His "amen" was the loudest after every special I sang at church, and I'm sure I heard his hands clapping above the crowd at each graduation. He's always been there to give me a hug, tell me he was proud of me, and remind me that he and Grandma loved me.

He's been more than just a voice of praise. He's been an example of a life lived in service to Christ and love for his family. I've joked with my wife that she should watch what Grandpa is like now because that is what I'll be like when I get older. The truth is, I have always been proud when someone tells me I take after my Grandpa Ross. He's been a model of the kind of man I want to be – one with strong conviction, the courage and boldness to share it, and a heart full of love.

Writing this dissertation has been the capstone to achieving one of my life's goals. I know Grandpa is proud and has been anxious for the day when I become Dr. Weaver. But reaching this point is minor in comparison to a much larger goal I've always had: to be a man who can do the same for my family that my Grandpa has done for ours. Grandpa has always pushed me. When I was a kid, it was on a John Deer tricycle he

rescued from the dump and restored for me. As an adult, he's pushed me to be the teacher, husband, father, and man God has always planned for me to be.

I want to dedicate this paper to the man who has pushed me. I love you, Grandpa.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Research identifies several grading practices considered ineffective for their negative impact on the accuracy of the report of student achievement they yield. The ability to positively impact student achievement is the definition of a teacher's sense of self-efficacy. This study examines the two factors to determine if a relationship exists between the grading practices of a teacher and their level of self-efficacy.

Teachers across the 20 counties that make-up southwest Missouri were surveyed using two instruments: the Grading Practices Inventory developed through research of best practices specifically for this study and the existing Teacher's Sense of Efficacy Scale developed by Dr. Anita Woolfolk-Hoy. The sample included teachers from kindergarten through grade 12, from districts ranging in size from less than 500 students to over 10,000 students, and with a broad range of years experience and degrees earned. The sample also included teachers who report single overall letter grades and those who report grades separated by individual standards or competencies.

Though results failed to demonstrate any relationship between the two factors, analysis of the results of each scale across several demographic factors produced several significant relationships between sub-groups of the demographic which helped to understand the nature of grading reform and served as the basis for recommendations. Recommendations included greater professional development for teachers about effective grading practices and further study into possible relationships between specific grading practices and efficacy levels of teachers.

CHAPTER ONE

INTRODUCTION

One of the duties of any teacher is the appraisal of student learning. Grading student work, and the summary of that work into a statement of overall progress, has been the subject of review and reform (Guskey, 1994; O'Conner, 2009; Brookhart, 2011). Questions raised about grades examine the purpose for grades, factors included in those grades, grade referents, and the calculation of final grades. Several of these grading practices, such as the use of grades as motivators (Guskey, 2001b, 2004), the inclusion of non-achievement factors in grades (O'Conner, 2009), norm-referencing instead of criterion referencing (Brookhart, 2011), and assigning zeros in a 100-point scale (Dueck, 2014) have been challenged for the detrimental effects on student learning and the accuracy of the report.

How a teacher views his or her capacity to influence student learning (the concept of teacher self-efficacy) can be influenced by their experiences that show student achievement and the stress and thought processes the teacher has about their work (Hoy, 2000; Brown, 2012). Teacher efficacy, a contextual application of Bandura's (1977) self-efficacy concept, has been linked to increased student achievement (Anderson, Greene, & Loewen, 1988; Moore & Esselman, 1992; Tschannen-Moran & Barr, 2004; Schumacher, 2009) and a more personal, less controlling style of interaction with students (Woolfolk & Hoy, 1990). Teacher efficacy has been shown to be a stronger influence on student achievement than other factors, such as socioeconomic status (Goddard, Hoy, & Hoy, 2004) and minority status (Watson, 1991).

Some current grading practices are challenged for the way they falsely inflate or deflate the achievement they show. Others are challenged because they are practices meant to influence student behavior instead of promote learning. Both student achievement and control in the classroom are related to the level of efficacy a teacher has. This study attempts to examine if there is a relationship between grading practices a teacher uses and the level of efficacy a teacher reports.

Problem Statement

Grading practices and a teacher's sense of efficacy appear to have several possible relationships. Higher levels of teacher efficacy are positively correlated to higher student achievement (Anderson, Greene, & Loewen 1988; Ross 1992; Goddard, Hoy, & Hoy, 2004**)**. The experience of seeing one's own students succeed builds teacher efficacy (Putman, 2012). However, specific grading practices such as norm referenced grading (Brookhart, 2011), the inclusion of non-achievement factors in grades (Guskey, 2006), and the use of extra credit (Welsh, DAgostino, & Kaniskan, 2013) are challenged for their inflation or distortion of student achievement. If teachers view student achievement that has been distorted, it may have an impact on the development of their efficacy.

Teacher efficacy and grading practices may connect when examining the management style of classrooms and the role grades may play in a teacher's style. Teacher efficacy levels are related to the management style and practices in a teacher's classroom in that teachers with low efficacy are more custodial with teachers attempting to assert control over their students (Woolfolk & Hoy, 1990). Grading practices such as the use of zeros and penalties for missed deadlines are practiced as a way of managing or influencing student behavior (Guskey, 2004).

As research continues in the area of grading practices, reform efforts proposed require fundamental change to a teacher's daily practice (Marzano, 2000; O'Conner, 2009; Brookhart, 2011). Changing a teacher's practice is often challenging work and teachers with higher levels of teacher efficacy are more willing to persist through change (Spark, 1988). Both accuracy in reporting student learning and building teacher efficacy are desirable in schools. Determining if a relationship exists between a teacher's grading practices and a teacher's sense of self-efficacy may assist teachers and administrators in planning for and implementing grading reforms and fostering greater efficacy among teachers.

Rationale/Purpose of Study

This study explored a teacher's grading practices and teacher efficacy to determine if a relationship existed. Accurate descriptions of student achievement are desired by parents, students and teachers as well as by local, state and federal government officials who link teacher performance evaluation to student achievement. In making the argument that some current grading practices may not result in accurate pictures of student learning, researchers are challenging a very basic yet important aspect to educational systems. The emphasis placed on student achievement and the necessity that achievement reports be accurate leads many schools and teachers to examine their grading practices. Sometimes, teachers are hesitant to adopt different grading practices because they are unfamiliar with alternatives and are more comfortable with their current practice.

Efficacy refers to an individual's belief in their ability to act in such a way to produce desired results (Bandura, 1977, 2000, 2006). Teacher efficacy then is the belief

a teacher has in his or her ability to perform tasks that positively impact students, especially student achievement (Ross, 1995; Tschannen-Moran, Woolfolk-Hoy, & Hoy, 1998; Weasmer & Woods, 1998; Dellinger, Bobbett, Oliver, & Ellett, 2008). Teachers who are highly efficacious have been shown to have a greater impact on student achievement (Goddard, Hoy, & Hoy, 2004). These teachers are also more resilient and likely to persist in the face of change (Cervone, 2000). Because of these relationships, it would be desirable to implement practices that build teacher efficacy because of the impact it would then have on student achievement. Identifying a relationship between teacher efficacy and grading practices may assist administrators and teachers interested in reform of grading practices by providing motivation and direction in planning such reforms.

Research Questions

This study sought to explore if a relationship between teacher efficacy levels and grading practices exists. Further, the study examined trends regarding the levels of effective grading practices and teacher efficacy across various demographics. The following research question guided the study: Is there a relationship between a teacher's grading practices and the level of teacher efficacy?

Subset Questions

1. What are the grading practices of a teacher?
2. What are the efficacy levels of teachers?
3. How do the efficacy levels and grading practices of teachers differ across gender, grade-level taught, district size, years of experience, and level of education?

Theoretical Framework

Examination of grading practices has been the subject of research conducted by several researchers including prominent names such as Marzano (2000, 2010), Guskey (1994, 2001, 2004, 2006, 2011, 2013), O'Conner (2009, 2011), and Brookhart (1994, 2011). The work of these researchers examined the underlying motivation behind specific practices as well as their impact on the accuracy of student achievement reported. Marzano (2000) identified five general purposes for grading. Grades are used to determine student retention or promotion, to rank and sort students for comparison, and to determine placement. Grades are also used to inform educators as they advise students about courses of study. Grading provides feedback both to the student about their learning and to teachers about the quality of their instruction. Frequently, grades are used to inform future instructional planning. Finally, grades are used as motivators for students.

Though multiple purposes for grades are given, the most commonly accepted is the use of grades to report student learning (Brookhart, 1994; Guskey, 1994). Given this purpose, specific practices are examined for their impact on the accuracy of final grades reported. The inclusion of non-achievement factors such as participation and attendance is questioned for distracting from the accuracy of actual student performance (Guskey, 2006). The manipulation of grades to influence behavior by penalizing student work for missing deadline or not following directions is questioned because it results in inaccurate performance levels (O'Conner, 2009; Dueck, 2014). The use of zeros in a percentage scale has a drastic effect when averaged with other marks for performance again resulting in distorted overall performance levels (Guskey, 2004; O'Conner, 2009)

The concept of efficacy originated in the work of Bandura (1977) who defined it and described four ways in which it could be developed in an individual. Mastery experiences refer to the experiences of an individual. The success or failure in those experiences directly leads to the sense of efficacy the individual has for doing that very action (Bandura, 1995). Vicarious experiences occur when an individual witnesses someone else conduct the action and translates their success or failure into their own ability to do the same (Bandura, 1997). Verbal persuasion refers to the feedback and expression of confidence in an individual's ability from people significant to the individual or whom the individual recognizes as having mastery (Bandura, 1997). Physiological and psychological state of the individual is described as a self-fulfilling prophecy; if an individual thinks they may fail, the belief creates stress and that stress can then impact the performance of the task (Bandura, 1997).

The concept has been applied to the field of education by several researchers who have shown relationships between efficacy levels and a teacher's approach to working with students in the classroom (Woolfolk & Hoy, 1990), classroom management (Woolfolk, 2001), and student achievement (Anderson, Greene, & Loewen, 1988; Moore & Esselman, 1992; Tschannen-Moran & Barr, 2004; Schumacher, 2009). Woolfolk-Hoy and Tschannen-Moran have spent considerable research refining the earliest definitions and measures of teacher efficacy to develop a commonly accepted evaluation tool today.

Limitations/Delimitations

Limitations.

This study examines school districts in southwest Missouri. Some of these districts have established grading policies in place while others give teachers their own

leeway to develop practices for grading. The difference in levels of teacher ownership over their own grading practices may influence their sense of how beneficial the practices are or their primary purpose in designing them.

The southwest region of Missouri mirrors the state in that a majority of districts are rural with less than 1,000 students in the district. The lack of larger districts in the region may limit the ability of the study to be generalized outside of the state.

An online survey instrument was used to collect data on individual teachers from those schools to identify their grading practices and levels of teacher efficacy. Surveys relied upon the honesty level of teachers completing them and their ability to self-analyze, which may have impacted the results. Because the survey items measured perception, their accuracy in defining those perceptions and the accuracy of individuals reporting those perceptions was a limitation.

Delimitations.

This study examines school districts in southwest Missouri. Some of these districts have established grading policies in place while others give teachers their own leeway to develop practices for grading. The difference in levels of teacher ownership over their own grading practices may influence their sense of how beneficial the practices are or their primary purpose in designing them.

Summary

Grading is an inescapable task associated with teaching. In the pursuit of accurate and reliable grades, several grading practices have been questioned by researchers for creating distortions in the way they report achievement. Overall student achievement and classroom styles have been linked to the levels of efficacy a teacher has. This study

sought to determine if any relationship existed between the grading practices of a teacher and that teacher's level of efficacy.

CHAPTER TWO

REVIEW OF LITERATURE

While there is considerable literature about grading practices and teacher efficacy separately, little is available that connects the two. In this chapter, each topic will be reviewed to establish separate frameworks in preparation for identifying relationships between them. The review begins with literature regarding grading practices by examining the history of grades before exploring how grades are used. Then, practices regarding types of grading references, composition of grades, assessment evidence used in determining grades, and calculation practices used in determining final grades will be reviewed. Next, a review of teacher efficacy literature looks at the origin and development of the theory of teacher efficacy, sources of developing teacher efficacy, the impacts of efficacy on teaching, and measures of teacher efficacy.

Grading

Grading student learning is a central function of teachers. Grades impact both teachers and students (Liu, 2007) and account for a considerable portion of a teacher's time (Stiggins, 1992). Teachers' grading practices are most often based on their personal beliefs instead of principles of evaluation and measurement (Brookhart, 1997; Allen & Lambating, 2001; Cox 2011). Often times, those beliefs are shaped by their own experiences as students (Lambating & Allen, 2002).

History of grading and reform

Grading practices and ideals have developed in schools and universities over many years. In the early 1700s, universities provided feedback regarding the quality of work to students without assigning actual grades (Marzano, 2000). In 1789, as enrollment

began increasing, Yale University began the practice of assigning students a score from a scale of four points to represent overall achievement (Tocci, 2010). Harvard University adopted a similar system by 1830 but soon moved to a 100-point scale (Tocci, 2010) divided into six divisions (Marzano, 2000). Included in both final grades at Harvard and Yale were how many times students attended classes or chapel services (Peabody, 1888). By 1850, the College of William and Mary began the practice of assigning percentage based-scores and in 1897, Mount Holyoke University began assigning letters grades to those divisions (Marzano, 2000).

In public schools, teachers in one-room school houses in the late 1800s listed which skills a student should be able to do and moved them on to the next grade when the student had mastered those skills (Guskey, 1994). Student grades based on their class performance were used to sort and order students. Students who did well were moved to the front of the room while students who did poorly were sat at the back. Reordering took place daily in some instances (Kaestle, 1983).

In 1908, the Boston public school system was the first to use straight percentages to report grades. Percentages of right or wrong answers on final exams were believed more accurate as they took the teacher judgment out of ranking students (Tocci, 2010). It was the early 1900s before public high schools began consistently adopting the practice of assigning percentages to students as marks of achievement. The use of percentages for reporting achievement came as a means of convenience in the face of increasing enrollment from compulsory attendance laws (Guskey, 2013) and gave grades a more official, objective appearance while easier to manage and report for large numbers of students (Guskey, 1994). Additionally, the need arose for grades to be reported in quick,

easily comparable means as universities, businesses, and even public school leaders themselves began looking at grades as a factor in admissions, hiring and promotion respectively (Schneider & Hutt, 2014). With high rates of transiency, educators also needed grades that could be universally recognized and transferred between schools (Synder, 1993). By 1918, teachers began reporting letters instead of percentages (Guskey, 1994). By the 1960's, letter grades were considered traditional and points or percentages were converted to letter grades for greater consistency (Burke, 1968). Grades during that time were additionally believed to be an efficient way to filter students into tracks for varying career options (Kliebard, 1987).

While many valued the apparent motivational aspect of rank associated with grades (Campbell, 1921), some educators questioned the use of competitive grading instead of a focus on learning (Schneider & Hutt, 2014). As early as 1912, a study conducted by Starch and Elliott was able to demonstrate considerable variation among teachers' grades of the same assignment pointing to inconsistent grading practices (Ciezk, Fitzgerald, & Rachor, 1996). In this study, teachers asked to grade the same essay yielded final grades ranging from 50 to 90 on a 100-point scale. Repeating their own study using math, teachers graded the same work with scores ranging from 28 to 95 on a 100-point scale (Guskey, 1994). Replicating the same study almost a century later yielded the similar results (Guskey, 2013). A study by Canady and Hotchkiss (1989) noted inconsistent practice resulting in drastically different grades between teachers within the same school and even between different periods for the same teacher. Analysis of these grading practices shows the variety of factors weighted and considered in determining a grade, such as number of items correct, difficulty of the exam, how the class performed,

and individual ability and effort, results in uncertain meaning for final grades assigned (Ciezk, Fitzgerald, & Rachor, 1996).

In an attempt to return a focus of grade reports back to learning, reformers in the 1930's advocated for the use of descriptive letters of individual progress instead of grade cards (Geyer, 1938); however, the volume of students and information to be compiled in light of the desire to easily communicate progress to entities outside the school and allow grades to remain portable prevented any such movement (Kvaraceaus, 1939). A study in 1971 sponsored by the National Educational Association revealed only 16 percent of elementary teachers and only 35 percent of secondary teachers believed letter grades appropriately reflected learning, however, teachers did not believe they could get away from issuing them (National Educational Association, 1971). At the university level, grade inflation became a greater concern as faculty evaluations were linked to student feedback (Healy & Rojstaczer, 2012) and as professors realized the impact of issuing a failing grade to a student impacted his/her draft eligibility status and not just his/her learning (Birnbaum, 1977). Developments in grading reform during the 1990's and 2000's seem to be driven by the move for standards-based instruction (Cox, 2011). Since the beginning of the 21st century, all states have some form of state-wide standards in place for mathematics, English, social studies, and science with most of those conducting state-wide assessments based on those standards (Wolfe, Viger, Jarvinen, & Linksman 2007). However, research continues to suggest that there is a gap between what is recommended and what is practiced in classrooms (Cox, 2011).

Purpose of grades

Grades serve many purposes and can have varied meanings for different groups. Grades can impact a teacher's expectations and self-perception (Stiggins, Frisbie & Griswold, 1989). Teachers see value in grades for judging student progress and for the impact it has on student motivation and achievement (Liu, 2007). However, the intended uses of grades varies among educators and there is little communication among colleagues about personal beliefs (Allen, 2005).

Marzano (2000) identified five general purposes for grading. Administratively, grades are used to determine student retention or promotion, to rank and sort students for comparison, and to determine placement. This is more often the purpose for teachers in inclusion classes (Grimes, 2010). Grades are also used to inform educators as they advise students about courses of study. Grading provides feedback both to the student about their learning but also for teachers about the quality of their instruction. Frequently, grades are used to inform future instructional planning. Finally, grades are used as motivators for students (Marzano, 2000).

The primary purpose for using grades according to teachers is to report academic progress (Cox, 2011). When asked to rank various purposes according the importance, 65 percent of middle school teachers identified reporting achievement against academic standards as the primary purpose (Grimes, 2010). Grading reformers such as Brookhart (1994) and Guskey (1994) also emphasize grades should be about communicating what students have learned, can do and should be able to do at their level. However, Brookhart (1994) acknowledges there is a gap between recommendation and teacher performance.

In addition to educators, parents have varied purposes for grades as well. Brookhart (1994) cautions that without a conversation with parents, cultural and personal expectations can greatly vary. Parents often interpret average differently, accepting lower grades than teachers (Waltman & Frisbie, 1994). Guskey (2002, 2006, 2009) states that parents value the detailed, prescriptive information of student performance evaluated according to detailed academic standards. However, a contradictory study indicated parents still value letter grades and information that allows them to compare their student to other children (Huntsinger & Jose, 2009). Similarly, college and university admissions have also indicated a desire for weighted grade-point averages that allow quick screening and comparison of applicants (Talley & Mohr, 1993).

Common to both parents and teachers is the use of grades as motivators for students. Twenty-one percent of middle school teachers identified motivation as the third most significant reason for grades. Use of grades as motivators is more common in higher grades and in English classes (Grimes, 2010). The belief that grades motivate students can be found in some research (Stiggins, Frisbie & Griswold, 1989; Brookhart, 1994); however, the use of grades to motivate is largely based on teacher and parent beliefs regarding schooling and grades (Cox, 2011). Grades sometimes serve the purpose of being a motivator for student academic behavior. As a result, grades are often manipulated as a consequence to missing deadlines, not following directions, or incomplete work (Guskey, 2004). Reformers argue against this practice philosophically from the standpoint that poor work ethic will naturally impact the level of a student's achievement (Dueck, 2014) and that the consequence of a manipulated grade is too far removed from the behavior to have the desired impact (O'Conner, 2009; Dueck, 2014).

Guskey (2001b, 2004) states that no study shows that negative grades motivate students. Instead, researchers more commonly see an inverse effect on student motivation as students become less and less likely to make an effort in the face of failing grades (Guskey, 2004). Students are instead more likely to disregard the grade or take it as a reflection of who they are (Selby & Murphy, 1992; Guskey, 1994; Ring & Reetz, 2000). Grades in this critical sense foster an idea in students that their mistakes are failures that are recorded and cannot always be overcome. This notion is contrary to prevailing ideas about human learning and development (Spady, 1991). Guskey (2001b) observes that students cheat the system by taking watered down courses to avoid failing grades.

Research in the area of grades and motivation revealed a similar detriment to learning when good grades are used as rewards. Students given grades as rewards achieve at lower levels than those who work without the reward (Kohn, 1993). Additionally, grades given for the purpose of recognizing or rewarding student performance are linked to reduced motivation, lesser desire for challenging tasks, and lower quality of thinking (Kohn, 1999). A Norwegian study involving 3,800 secondary students showed that students receiving inflated grades at or above their actual performance level were less likely to study and received lower achievement marks (Bonesronning, 2004). Teacher use of grades as motivation is rarely based on scholarly research and more often on what teachers have experienced themselves or feel to be right (O'Conner, 2009). Reformers recommend presenting students with specific information regarding expected and actual performance to develop students into self-motivated learners (Brookhart, 2011).

Researchers have maintained that grades are not necessary for learning (Frisbie & Waltman, 1992). Assigning grades places teachers in a judgmental role and often conflicts with other roles for a teacher (Bishop, 1992). Descriptive feedback aligned to clear learning intentions is necessary for student learning (Brookhart, 2008). Some teachers report greater student effort and work completion when students receive meaningful feedback rather than just grades (O'Conner, 2009).

Grade referents

Varying purposes for grades have yielded various referents for grades. These referents include grades based on academic standards as well as those comparing present performance to the overall performance of a class or group. Grades typically are assigned to reference a pre-determined distribution, an objective standard for performance, or effort and performance gains (Marzano, 2000). Barnes (1997) categorizes these as norm-referenced and criterion-referenced grades.

Norm-referenced grades report where a student's performance falls in comparison to other students (Barnes, 1997). At the university level, the University of Missouri was the first to begin the practice of curving grades in 1908 by fitting all student scores into a standard bell curve in response to greatly varied failure rates among students from different departments (Tocci, 2010). Research from the time about student intelligence on normed tests, began the practice in the 1930s of curving grades to fit a standard bell curve. Normed tests support the sorting and ranking purpose of grades (Barnes, 1997).

The validity of normed grades has been the subject of much controversy (Guskey, 1994). Normed grades are believed detrimental to learning (Brookhart, 2011) especially to students with disabilities (Pollaway, Epstein, Bursuck, Roderique, McConeghy, &

Jayanthi, 1994). Normed grades shift the focus from learning for mastery to outperforming peers (Guskey, 2001a) creating the idea that for a student to help a peer actually hurts himself (Guskey, 2001b). Guskey (2011) argues that normed grades are not valid because the normal bell curve assumes no interventions and since teaching is an intervention, the bell curve can no longer be expected to apply.

Criterion-referenced grades report how a student has performed against an objective set of standards or level of performance (Barnes, 1997). Criterion-referenced grades require clearly identified standards for students to meet (Brookhart, 2011). This method supports the purpose of grades being measures of achievement and learning and is considered more rigorous than normed grades (Barnes, 1997). Criterion-referenced grades for the purpose of measuring achievement are the most widely recommended by researchers and reformers (Guskey & Bailey, 2001; Wiggins & McTighe, 2006; O'Conner, 2009, 2011;Marzano, 2010; Brookhart, 2011). The value lies in sharing explicit targets for students allowing them to take ownership and monitor their own learning (Brookhart, 2011). Teachers must first dissect standards to uncover and organize each component of standards (Guskey, 2005), however their use in criterion-referenced grading has been linked to learning gains a full standard deviation above traditional teaching (Guskey, 1980, 2007a).

Grades that reference improvement and effort are not recommended as stand-alone grades, but instead as a supplement to criterion-referenced grades (Guskey, 2006; Guskey, Jung 2009). These grades are often referred to as value-added grades because they measure the impact of learning that has taken place (Guskey, 2006). Grimes' (2010) study of teacher beliefs about grading revealed that 98 percent of math teachers, 95

percent of social studies teachers, 87 percent of science teachers and 87 percent of English teachers report progress or improvement grades. Teachers consider progress grades in an effort to maintain or improve student motivation, esteem and avoid social consequences of low grades (Guskey 2006). However, reformers and students alike discourage the common practice of including those grades as a factor in a single overall grade (Bursuck, Mink, & Olson, 1999; Guskey, 2006; Brookhart, 2011). The practice can also be harmful to special education students where improvement grades as factors in final grades can mask deficiencies (Bursuck, Polloway, Plante, Epstein, Jayanthi, & McConeghy, 1996).

Factors included in grades

Teachers can consider a wide array of factors in determining a final grade. In addition to academic achievement on assessments, assignments, and projects, teachers may consider non-achievement factors such as work completion, timeliness, effort, attitude, attendance and behavior (Guskey, 2006). While some place value on these factors (Marzano, 2000), others caution against including non-achievement factors in grades (O'Conner, 2009).

In a survey of elementary and secondary teachers, Ciezk, Fitzgerald, and Rachor (1996) found 61 percent of teachers using non-achievement factors in their grades. These include effort, teamwork, and adjustments made for difficulty of task. Marzano (2000) also found 35 percent of teachers from kindergarten to high school using effort as a factor in final grades. Even when looking at measures of academic achievement such as test or assignments, teachers considered student ability, overall performance of the class and the difficulty level of the task with the percentage or number of items correct (Ciezk,

Fitzgerald, & Rachor, 1996). Non-academic extra credit is also often a factor (Liu, 2007; Welsh, D'Agostino, & Kaniskan, 2013) though often few students take advantage of it and those that do are ones who are already academically performing well (Harrison, Meister, & LeFevre, 2011).

Guskey (2006) notes discrepancy among teachers as a result of mixed or unclear purposes for grading. While many teachers identify reporting academic achievement as their primary purpose, few include mastery of learning objectives as the major component in a final grade (Grimes, 2010). In a study by Ciezk, Fitzgerald, and Rachor (1996), researchers found a high level of variability among the number of marks per term teachers included in grades. The results from a survey of 113 teachers shown an average of 24.3 assignments per term with a SD of 17.9. Further, the researchers found teachers who include more marks per term are 1.04 times more likely to also include non-achievement factors.

Factors such as attendance, behavior, work ethic and participation are important for a student to receive feedback and are valued by parents, teachers and employers (Marzano, 2000). However, researchers recommend reporting those factors separate of an academic grade (Canady & Hotchkiss, 1989; O'Conner, 2009; Grimes, 2010). Cross and Frary (cited in Liu, 2007) found a majority of teachers and students agree with the idea of reporting non-achievement factors separately. Some teachers have reported students demonstrating more attention to those factors when they receive separate grades and feedback about them (Guskey & Jung, 2009). In addition to the apparent benefits to students, O'Conner (2009) cautions teachers that courts have challenged and overturned grades based on non-achievement factors.

Assessment practices in grading

The assessments including tests, projects and assignments that grades are based on play an important role in the measurement and reporting of learning (O'Conner, 2009; Welsh, D'Agostino, & Kaniskan 2013). Ciezk, Fitzgerald, & Rachor (1996), based on a study of 113 mid-western teachers, found that 60 percent of teachers prefer to write their own major assessments and 80 percent of teachers prefer to write their own minor assessments. The researchers also found that 96.9 percent of the teachers in their first five years of teaching write their own assessments. However, research indicates there is little training for teachers in how to design quality assessments aligned to standards resulting in teacher bias in the assessments (Zhang & Burry-Stock, 2003; Guskey, 2006; Wolfe, Viger, Jarvinen, & Linksman, 2007).

Measurement researchers and grade reformers caution teachers to be aware of the limitations and biases of an assessment (Stiggins, 1992). Assessments should be aligned clearly to standards (O'Conner, 2009) and should be accompanied with descriptive feedback to be most effective in continuing learning (Brookhart, 2008). Rundquist (2012) advocates including student voice in the form of reflection on what they have learned to deepen the value of assessments. Knowing that a degree of error exists in all assessments, teachers should base their final grades on multiple assessments (Welsh, D'Agostino, & Kaniskan 2013). Additionally, many standards are complex requiring multiple measures to accurately assess various aspects adequately (Brookhart, 2011). Common sources may include tests and quizzes, portfolios, projects, laboratory reports, student notebooks or journals, and teacher observations (Guskey, 2006). Group grades are often discouraged because they distort individual achievement (Brookhart, 2011).

Administrators generally consider standardized assessments, such as state or district tests and end-of-course exams, as more valid than teachers who are more likely to consider their own observations and student participation as valid tools (Guskey, 2007b). Educators must select assessment tools that are of appropriate format for the skill or concept being assessed (Stiggins, 1992), but generally, the more assessments included, the more accurate the picture of student achievement (Brookhart, 2011).

Researchers have identified diversity among teachers in terms of both the frequency of assessments and number of assessments given during a marking period (Ciezk, Fitzgerald, & Rachor, 1996). Liu (2007) found 43 percent of teachers surveyed provide students the chance to retake a failed assessment while 37 percent believed a second chance invalidated the results. There remains little consensus among teachers and researchers as to how many measures to take and how to combine and interpret the combined results of those assessments (Henderson-Montero, Julian, & Yen, 2003).

Calculation of final grades

Stemming from attempts to increase objectivity of assessments and make processing of assessment information for large numbers of students easier, teachers have relied on using points or percentages of items correct as a calculation method for grades since the turn of the twentieth century (Marzano, 2000; Tocci, 2010). However, researchers point out that numbers and percentages can still include degrees of subjectivity and it is the measurement tool itself that matters (O'Conner, 2009; Guskey, 2013). The use of points and percentages is more reliable when considering a large number of items about the same topic or concept; however, this is often not the case in classroom grades (Marzano, 2000; Brookhart, 2011). Using percentages allows

instructors to set cut scores for proficiency. High cut-off scores are often seen as more rigorous, however again it is the measurement tool itself that has more to do with the level of rigor in the assessment (Guskey, 2001a). Higher cut scores skew toward the negative with many scales including 60 degrees of failure and 40 degrees of success which is difficult to accurately and consistently measure between teachers (Guskey, 1994, 2013).

Scores can further be skewed from reflecting academic achievement accurately when they are modified by teacher penalties (Guskey, 2004). Common punitive practices that distort measures of academic performance are reducing scores by a given percentage for being submitted past the deadline or assigning a zero to missing or late work (Dueck, 2014). Penalties like this not only distort the true picture of performance, but can decrease student motivation to complete those tasks once the deadline has passed (O'Conner, 2009). As an alternative, Guskey (2001b, 2004) suggests assigning incompletes or reporting those behaviors separately.

Another concern with assigning zeros as scores is the impact they have when averaged together with other evidence to reach a final grade (Guskey, 2001b, 2004; O'Conner, 2009; Dueck, 2014). While using the mean of a set of scores is the most common practice, other methods such as using the median or mode for the same set of scores often produce more accurate pictures of student achievement, especially in the presences of zeros as scores (Marzano, 2000; O'Conner, 2009). Reformers hypothesize that one reason zeros are assigned is as a punitive tool to impact the average of a student (Guskey, 2004).

Further considerations for calculating scores include giving greater weight to more recent evidence. A simple mean holds students accountable for mistakes made early in their learning while other methods of calculation allow the most recent evidence of possible mastery to be clear (O'Conner, 2009; Brookhart, 2011). Additionally, considering performance against an analytic rubric to identify discreet skills a student has or has not mastered is preferable (Brookhart, 2011). The smaller scale of such rubrics is less influenced by zeros or averages (O'Conner, 2009) and has been shown to have a stronger correlation with student achievement on standardized tests than points and percentages (Wright & Wise, 1988).

Validity and variability of grades

In a study of grade cards issued at different grade-levels in Wisconsin, Freedman and Frisbie (1995) found that 100 percent of high schools report letter grades only. At the middle school level, only 10 percent of schools reported grades by individual standard while at the elementary level, 61 percent of schools were reporting by standard instead of a single, overall letter. A national survey of middle school principals from another study by MacIver (1990) found 99 percent of schools reporting single letter grades with 50 percent reporting conduct and 27 percent reporting effort separately. Marzano (2000) cites research from Robinson and Craver (1989) that shows letter grades increasingly more common in higher grades while skill checklists were found in mostly elementary settings.

Providing a single letter grade as a summary of all learning is a difficult and often challenged practice. A single letter cannot summarize the complex information about a student's learning including progress across many complex standards and various work

behaviors in any meaningful way and results in a hodgepodge grade (Guskey, 2006, 2011). Single grade summaries of student learning have been questioned for their reliability and uncertain meaning (Friedman & Frisbee, 1993; Brookhart, 1994).

The unreliability of grades has been a concern for over a century. Researchers Starch and Elliott conducted a study in 1912 asking several teachers to each grade the same writing sample and the resulting grades varied over 40 points among participants. A similar study of teachers of mathematics a year later resulted in even greater discrepancy (Starch & Elliott, 1913). The findings were replicated producing congruent results almost 100 years later. A group of 90 high school English teachers were asked to grade the same work after receiving 20 hours of training specific to assessment of writing and the final grades they issued varied by 46 points (Brimi, 2011).

Much of the unreliability of grades stems from the highly individualized and diverse grading policies teachers develop (Liu, 2007; Cox, 2011). Practices may vary among teachers from different grade levels (Randell & Engelhard, 2009) and between different periods taught by the same teacher (Canady & Hotchkiss, 1989). Variations among teachers include assigning different weight values to the same evidence (Marzano, 2000) and considering different factors such as progress over performance (Friedman & Frisbie, 1995). These differences exist despite most districts having a written policy of some form or another in place for grade assignment (Polloway, Epstein, Bursuck, Roderick, McConeghy, & Jayanthi, 1994). Stiggins, Frisbie and Griswold (1989) suggest that these differences may exist because recommended measurement practices are too difficult to be realistic for teachers or teachers ignore the recommendations because of their own lack of knowledge.

Teachers are prone to modify grades individually for students as they consider unique student circumstances (Guskey, 2006; Guskey & Jung, 2009; Gordon & Fay, 2010). Grades for special education students are most commonly subject to modification including modification of grading criteria or scale and modification of the final grade (Munk & Bursuck, 1998). One study by Polloway, et al, (1994) showed policies providing for the modification of grade of a special education student are in place in 64 percent of districts. Another study by researchers Valdes, Williamson, and Wagner (1990) found that only 64 percent of special education student in general education classes were held to the same standards as other students in the same course. The same study found course standards modified for 74 percent of special education classes.

Marzano (2000) notes non-educators hold an assumption that grade inflation is common, perhaps as a result of the highly individualized grading practices of teachers. Ciezk, Fitzgerald, and Rachor (1996) note a bias in teacher practices toward factors that produce higher grades such as dropping outlying low scores without also dropping outlying high scores. Marzano (2000) cites a study from the United States Department of Education's Office of Educational Research and Improvement finding a high degree of grade inflation from a national sample of eighth grade students where the free and reduced lunch rate was above 75 percent. In the study, the level of work in reading needed to earn grades of C or D in low poverty schools were show inflated to an A in high poverty schools. In the same way, D grades in math from low poverty schools were inflated to A's in high poverty schools. Another study sponsored by the College Board cited by Marzano (2000) reported a 9 percent increase in A-level GPAs while at the same time SAT average scores fell 13 points in the verbal section and 1 point in math. Grades

were shown to have a low correlation with standardized state assessments as a result of varied grading practices. Teachers' grades were most likely to converge in math and with scores at the meets expectations level (Welsh, D'Agostino, & Kaniskan, 2013). Teachers are more likely to attribute the low correlation to the assessments while policy makers blame variability in teacher grading practices (Guskey, 2006).

Students view many teacher practices as unfair. Bursuck, Munk & Olson (1999) found 86 percent of students believed grading for improvement was unfair and 56 percent believed raising grades for effort was unfair. Addressing modification of grading standards for special education students, 81 percent of students did not agree with reducing the amount of learning needed to earn a grade, 87 percent were against modifying the grade scale, and 95 percent disagreed with passing a student no matter what. Equally high percentages of disapproval came from special education students in the study. At the college level, Gordon & Fay's (2010) research indicates students view grades from instructors who provided learning assistance as more fair than those using grading practices to modify grades after the fact.

Efficacy

The concept of self-efficacy originates with the work of Alfred Bandura (1977). Bandura explains efficacy as a person's belief in his or her own ability to successfully conduct an action aimed at achieving a result. Efficacy is a completely cognitive function, assessing a person's own abilities (Bong & Clark, 1999). This belief alone does not make the action possible, but drives an individual to make the attempt (Pajares, 1996) and can be used as a predictor of success (Bandura, 1997).

Efficacy is distinguished from several other similar constructs such as self-esteem, self-concept, confidence, and locus of control. Self-esteem differs from self-efficacy in that esteem is the sense of overall value an individual places on him or herself while efficacy is specific to a given task or action. In this way, an individual may have high efficacy, but low overall esteem or vice versa (Bandura, 1977). Self-concept represents general feelings of ability where again efficacy is linked to a specific action and therefore can be used as a predictor of success (Bandura, 1997). Confidence differs from self-efficacy in that confidence generally describes a constant state of certainty in one's own ability while self-efficacy may differ from one circumstance or task to another (Glidewell & Livert, 1992). Most similar is the idea of locus of control. Locus of control refers to an individual's sense of whether or not something can be influenced by individual action. Self-efficacy then is the belief in one's own ability to conduct that action well (Bandura 2006).

Sources of Efficacy

Efficacy can be built from one of four sources (Bandura, 1977). Mastery experiences are the most powerful in developing efficacy. Mastery experiences refer to the experiences of an individual. The success or failure in those experiences directly lead to the sense of efficacy the individual has for doing that very action (Bandura, 1995). For example, mothers who have had the opportunity to care for infants prior to becoming a mother demonstrate high levels of efficacy in their ability to care for an infant than those who did not have similar experiences. Those mothers who had cared for infants before had increased efficacy as a result of their own personal successful experiences (Froman & Owen, 1990). In the same way, students who are able to complete assignments at their

instructional level and get the answer right demonstrate high levels of efficacy because of their own experience where students presented material above their instructional level who fail in their attempt have lower efficacy levels because of their experience (Margolis & McCabe, 2006).

The second most influential are vicarious experiences. Here, an individual witnesses someone else conduct the action and translates their success or failure into their own ability to do the same (Bandura, 1997). Downes (1993) demonstrates the power of vicarious experiences by studying practicum students whose supervising teachers used technology in their instruction versus those whose supervising teachers did not. Results showed greater levels of efficacy in those practicum students who saw technology integration modeled. Neck & Manz (1992) identified mental rehearsal of a task as a means of vicarious experience since the individual is not actually conducting the action but is watching the imagined experience of themselves. Applying this approach, Wang, Ertmer, and Newby (2004) demonstrated that preservice teachers who mentally rehearsed their use of classroom technology prior to actually teaching a lesson showed higher levels of efficacy than those who did not. Vicarious experiences are impacted by how the individual judges themselves against the person they witness. If the other person is viewed as someone on par with the individual's own abilities, then seeing him or her fail can have a negative impact on the individual's sense of efficacy. However, if the individual judges him/herself as superior to the other person, his/her sense of efficacy remains unchanged by the failures of the other (Brown & Inouye, 1978).

A third source of efficacy comes from verbal persuasion. This refers to the feedback and expression of confidence in an individual's ability from people significant

to the individual or whom the individual recognizes as having mastery (Bandura, 1997). Verbal persuasion in the form of assurances and encouragement from a significant other has been shown to increase the efficacy of new mothers (Reece, 1993). Weaker still than the impact of vicarious experiences, confidence expressed will be dismissed if the individual attempts the action and fails (Zeldin & Pajares, 1997). Students who fail after the verbal persuasion of their teachers that they can be successful may then in the future disregard their teacher's encouragement (Margolis & McCabe, 2006).

The final contributor to the development of efficacy, and the weakest of all, is the physiological and psychological state of the individual (Bandura, 1977, 1997). Bandura (1997) describes the impact of this source as a self-fulfilling prophecy; if an individual thinks they may fail, the belief creates stress and that stress can then impact the performance of the task. Further, stress caused by other factors can impact the performance. Reese and Harkless (1998) note the effect of stress as a contributing factor to lower rates of efficacy among new mothers. Bandura (1982) explains this impact occurs because of the physiological reaction individuals experience during stress is often viewed by the individual as a weakness or vulnerability.

Impacts of Efficacy

An individual's sense of efficacy influences that individual's choice in goals, the level of effort put into reaching those goals, and the level of persistence in the face of difficulty (Bandura, 1997). High levels of efficacy lead to greater commitment to goals and higher personal expectation (Bandura, 1997). High efficacy also contributes to an individual's sense of peace during times of resistance (Pajares, 1996). On the contrary, individuals with a low sense of efficacy tend to believe tasks are more difficult and then

become more stressed about the task (Pajares, 1996). When individuals believe they are incapable of completing the task well, they are less likely to continue to attempt when it becomes difficult (Bandura, 2000).

Bandura states that personal efficacy levels impact the mental models a person holds. Those mental models in turn impact performance. Individuals who imagine success tend to focus on things that go well, while those who imagine failure perseverate on things that go wrong (Bandura, 1995). The higher the sense of efficacy, the more like an individual is to look at him or herself as the cause of either success or failure (Bandura, 1997). Research has shown that the knowledge and skills an individual possesses are alone unreliable predictors of future success because of the strong impact personal beliefs about those knowledge and skills can have on actual performance (Pajares, 1996). Bandura (2000) suggests that programs designed to specifically build efficacy as part of learning a new task will result in better overall performance.

Teacher Efficacy

Bandura's concept of self-efficacy has been applied by many researchers to education and used to define teacher-efficacy. Teacher efficacy then is the belief a teacher has in his or her ability to perform tasks that positively impact students, especially student achievement (Ross, 1995; Tschannen-Moran, Woolfolk-Hoy, & Hoy, 1998; Weasmer & Woods, 1998; Dellinger, Bobbett, Oliver, & Ellett, 2008). Teacher efficacy is often divided into two similar ideas. General teacher efficacy (GTE) refers to the belief that educators as a whole can have a positive impact on students while personal teacher efficacy (PTE) refers to the faith a teacher has in themselves to impact a student (Ashton & Webb, 1986; Woolfolk, 2001).

Sources of Teacher Efficacy

Teacher efficacy can be developed in the same ways as general self-efficacy: mastery experiences, vicarious experiences, verbal persuasion, and physiological states (Tschannen-Moran, Woolfolk-Hoy, & Hoy, 1998). Mastery experiences that develop teacher efficacy come from the length and type of teaching experience, preparation program, professional development, and reflection. Each source of teacher efficacy is noted to have the same level of impact as the corresponding source for efficacy (Tschannen-Moran & Woolfolk-Hoy, 2006).

Mastery experiences are the greatest factor in the development of a practicing teacher's sense of efficacy (Tschannen-Moran & Woolfolk-Hoy, 2006). Teachers attempt a new strategy, see it work and the impact on students, and are reinforced resulting in higher levels of efficacy (Hoy, 2000). The greater the years of experience have an overall greater sense of efficacy with specific regard to instructional practices, classroom management, and student engagement. This may be a result of actual experience and the burnout of teachers with lower efficacy who leave the profession earlier (Putman, 2012). Not just the actual experience, but reflection on those experiences also improves efficacy as teachers refine and develop a deeper knowledge of content (Sibbald, 2008). Years of experience influence a teacher's personal sense of efficacy. Efficacy is generally observed to rise throughout teacher preparation programs but can sharply decline during the first year as a result of increased stress. Efficacy again rises with additional year of experience (Hoy & Spero, 2005). A teacher's personal sense of efficacy is further influenced by teaching assignment or content (Blazevski, 2006) and grade level (Marachi, Gheen, & Midgley, 2000).

Research has been done in the early formation of teacher efficacy through preparatory programs. Bandura (1993) states that efficacy is easiest to build in early developmental learning stages. Teacher training programs for master's or other advanced degrees show higher levels of efficacy in teachers than those with just bachelor's or alternative certification (Schwartz, 2010). Informal pre-service experiences such as camp counselors and child-care supervisors have been shown to improve efficacy specific to student engagement (Tuchman, 2010). Strong induction programs when teachers begin have also been shown to improve teacher efficacy (Collins, 2006). Professional learning experiences during teaching also have considerable impact on a teacher's personal sense of efficacy (Swars, 2005). The most effective professional development models for improving teacher efficacy include critical thinking and reflection (Hanson, 2002) or a personal experience supported with follow-up coaching (Tschannen-Moran & McMaster, 2009).

Teacher efficacy can be built through the vicarious experience of colleagues. As teachers hear about or witness a successful implementation by a colleague they view as effective, their own personal teaching efficacy is improved (Hoy, 2000). Likewise, observation of a poor performance can decrease an observer's level of efficacy (Hoy & Spero, 2005). Modeling has been shown an effective instructional method for teacher training and professional development (Wang, Ertmer, & Newby, 2004). The degree to which the observer identifies with the colleague they are observing affects the degree of impact on the observer's level of efficacy (Hoy & Spero, 2005).

Verbal persuasion in the form of positive feedback or encouragement from colleagues and evaluators builds teacher efficacy (Hoy, 2000). There is also speculation

that highly collaborative environments contribute to raised collective efficacy, or the sense of efficacy teachers in a building holds for their building (Hoy & Sabo, 1998). These environments often provide positive support and clear vision that improve the sense of efficacy (Jhanke, 2010). Increased collective efficacy in turn is linked to higher personal efficacy among individual teachers in the building (Kurt, 2009). Research also points to a positive relationship with a teacher's supervisor, the formal feedback and positive verbal support of the supervisor (Maginnis, 2009), the responsiveness of the supervisor (Newman, Rutter & Smith, 1989), and the principal's instructional leadership (Calik, Sezgin, Kavgaci & Kilinc, 2012) as factors that improve personal teacher efficacy.

Physiological and psychological factors can also impact a teacher's sense of efficacy. Those impacts that affect efficacy in teachers include feeling unprepared to address classroom management issues or parents, increased bureaucratic requirements, and depersonalization of the work environment. These factors result in anxiety and stress that deplete efficacy levels (Brown, 2012; Gillett & Urbanski, 2004).

Impacts of Teacher Efficacy

Teacher efficacy is an important factor in educational change, teacher burnout, learning environments, instructional practices, classroom management, and relationships with students and parents. Teacher efficacy serves as a factor in predicting student achievement and teacher practice. Despite many findings of the significance of teacher efficacy in research, little about it makes its way into field journals designed for teachers themselves (Alcock, 2003).

Teacher efficacy has a significant impact on the classroom environment and instructional practices of a teacher. Teachers exhibiting high teacher efficacy often have more self-directed classrooms (Woolfolk, 2001), plan more student-centered instruction (Allinder, 1994) and hold themselves and students to higher and more challenging standards (Ross, 1995). They are more likely to work through difficult content with students (Woolfolk, Rosoff, & Hoy, 1990) and encourage students to see themselves as learners (Ross, 1994).

Teachers with high levels of teacher efficacy are more personal in their approach to the classroom (Woolfolk & Hoy, 1990). Those teachers are likely to praise correct answers, persist in the face of incorrect answers, and remain adaptable and responsive to student needs during small group instruction (Gibson & Dembo, 1984). Teachers with low levels of teacher efficacy are more custodial in their approach to the classroom (Woolfolk & Hoy, 1990). Those teachers tend more to criticize and fixate on incorrect responses as well as to be more easily frustrated by distractions during small group instruction (Gibson & Dembo, 1984). Overall, teacher efficacy has been shown to have a significant, positive correlation on ratings of instructional quality by both students and teachers (Holxberger, Philipp, & Kunter, 2013).

Teacher efficacy also impacts classroom management. Teachers with higher levels of efficacy are more willing to work with difficult students for a longer time (Woolfolk, 2001). Teacher efficacy levels have been shown a valid predictor of whether teachers will handle a discipline problem themselves or write an office discipline referral (Hughes, Barker, Kemenoff, & Hart, 1993). The higher the level of teacher efficacy, the more likely a teacher will focus on instruction and less on controlling students (Yeo, Ang,

Chong, Huan, & Quek, 2008). Meanwhile, teachers with low teacher efficacy are less likely to recover from setbacks with students and tend to use external motivators and punishments (Bandura, 1993).

Teacher efficacy has a greater impact on student achievement than socioeconomic status (Goddard, Hoy, & Hoy, 2004) or minority status (Watson, 1991). In a study of 561 elementary schools, teacher efficacy levels were shown to significantly correlate to performance in math ($r = .547$, $p<.01$) and reading ($r = .436$, $p<.01$) (Schumacher, 2009). Another study reveals the same connection with 8th grade math, writing and English assessments (Tschannen-Moran & Barr, 2004). Teacher efficacy has been shown an accurate predictor of student achievement on the Iowa Test of Basic Skills (Moore & Esselman, 1992), the Canadian Achievement Test (Anderson, Greene, & Loewen, 1988), and the Ontario Assessment Instrument Pool (Ross, 1992). Further, teachers with high levels of efficacy are more willing to accept responsibility for assessment results and continue working through problems demonstrated by those results (Ross, 1995).

Teacher relationships with parents and students are impacted by a teacher's level of efficacy. Higher levels of efficacy mean teachers are more willing to work with students and are less likely to make referrals to special education (Hoy, 2000). Teachers with high teacher efficacy have been shown more successful academically with students of all ethnic backgrounds (Tucker, Porter, Reinke, Herman, Ivery, Mack, & Jack, 2005). Those teachers also generally have a better response from parents and seek to involve them more often (Morris, 1995).

A teacher's level of efficacy has been connected to students' levels of efficacy as well. Because teachers are considered a credible person in a student's life, their feedback

provides verbal persuasion that impacts a student's level of efficacy (Siegle & McCoach, 2007). Teachers with low efficacy who rely on punishment and reward to maintain behavior undermine a student's sense of efficacy (Bandura, 1993), while teachers with high efficacy who involve parents can have a positive impact on a student's sense of efficacy (Viaderi, 2005).

Teacher efficacy plays a role in educational change and is repeatedly shown to be a factor in a teacher's initiation of change (Cervone, 2000; Weasmer & Woods, 1998). Spark (1988) identifies teacher efficacy as one of three elements necessary for teachers to accept and implement a change initiative along with acceptance of the method's underlying philosophy and perception of the method's costs. Teachers must believe they can first have a positive impact before the change can occur (Enderlin-Lampe, 2002). Teachers with high levels of efficacy are more ready to accept change and therefore are less likely to reach burnout (Evers, Brouwers & Tomie, 2002). High teacher efficacy is correlated with higher levels of professional commitment (Coladarci, 1992). Low levels of efficacy are associated with teachers feeling greater emotional and physiological stress (Jex, 1999). Lower levels of efficacy is associated with increased rates of teacher burnout because the stress from lower levels of mastery, anticipation of behavior problems, and lower student achievement create a cycle that leads to burnout (Skaalvik & Skaalvik, 2007).

Measuring Teacher Efficacy

As the concept has been studied, several researchers have worked to develop measures of teacher efficacy. Teacher efficacy is a difficult concept to measure since a teacher's tasks are so varied making it possible to feel more or less efficacious depending

on the specific aspect of teaching (Bandura, 1997). Bandura (2006) has offered some advice to researchers attempting to develop measures, namely in the choice of language to help identify perceived capability which relates to efficacy instead of intention to act which does not.

Researchers at the Rand Corporation developed the earliest measure of teacher efficacy. It was a simple two-item survey that combined to provide a single efficacy score. The two items were "When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment," and "If I really try hard, I can get through to even the most difficult or unmotivated students" (Armor, Conroy-Oseguera, Cox, King, McDonnell, Pascal, & Zellman, 1976, p. 153). The RAND Corporation survey defined the two dimensions of teacher efficacy and remained the accepted measure of teacher efficacy until other researchers questioned the validity of tool and developed their own measures (Ashton & Webb, 1982, 1986; Gibson & Dembo, 1984). Researchers Ashton and Webb (1982, 1986) expanded the two questions from the RAND survey to develop two separate scales. From expanding the first item, they developed a scale for general teacher efficacy, and, from expanding the second, they developed a scale for personal teacher efficacy (Ashton & Webb, 1986). Expanding on Ashton & Webb's work, Gibson & Dembo (1984) developed the Teacher Efficacy Scale as a way to measure both dimensions of teacher efficacy and observable behaviors. In the development of their scale, Gibson & Dembo validated the two dimensions of teacher efficacy as being personal teaching efficacy referring to a teacher's personal belief in their own ability to influence student achievement in their classrooms and general teaching efficacy referring to the belief of

teachers that educators in general can positively impact students even in the face of family background, social environment, and parental influence. Gibson & Dembo further began identifying observable behaviors that corresponded to various levels of efficacy finding that teachers with higher efficacy spent more time with whole group instruction and more personal time planning than did peers with lower efficacy. Guskey and Passaro (1994) raised concern over the validity based on incongruent language that could confuse participant responses. Researchers Tschannen-Moran, Woolfolk-Hoy, and Hoy (1998) developed a scale further capturing teacher efficacy across three sub-domains: efficacy in student achievement, efficacy in instructional strategies, and efficacy in classroom management. Their scale, named the Teachers' Sense of Efficacy Scale (TSES), was refined and reduced from over 50 items to a long form with 24 items or a short form with 12 items. The development of this scale represents the currently most accepted tool for measuring teacher efficacy accurately and reliability.

Relationships Between Grading Practices and Teacher Efficacy

While there is an absence of direct research connecting grading practices and teacher efficacy specifically, there are several areas where research suggests a possible relationship between the two. Levels of self-efficacy may influence a teacher's choice in how grading practices are established. Several grading practices such as norm-referencing (Brookhart, 2011), inclusion of non-achievement factors (Guskey, 2006), and modification of grades based on non-academic penalties (Guskey, 2004) can all distort the true level of achievement. Through mastery experiences, teachers build their sense of efficacy when they are able to work with students and see the results of those efforts in student achievement (Hoy, 2000). If those achievement results are skewed showing

lesser achievement than is deserved, the development of a teacher's sense of efficacy may be influenced. Researchers have also noted frequently how inconsistent a teacher's grading practices may be with other teachers (Canady & Hotchkiss, 1989; Friedman & Frisbie, 1995; Marzano, 2000; Randell & Engelhard, 2009). Affirmations from peers as verbal persuasion is another method for developing a teacher's sense of efficacy (Hoy, 2000). Variance in teacher practice may keep peers from affirming different practices hindering the growth of efficacy.

A case study of five teachers who implemented a grading reform moving to a criterion-referenced, non-graded system showed the motivation for this change in each case was directly related to the desire in each teacher to put students in a position to understand their grades better and begin self-regulation of their learning (Percell, 2014). Other research into efficacy shows teachers with greater efficacy are more likely to invest in students academically and work actively to build a student's capacity to learn (Ross, 1994; Woolfolk, 2001, Yeo, et al, 2008). These findings hint at a possible relationship between teacher efficacy and the grading practices a teacher might adopt.

Research shows teachers with low efficacy are more likely to establish practices that control students (Woolfolk & Hoy, 1990). Several grading practices such as adding penalties to grades for missed deadlines and the inclusion of behaviors in a final grade are recognized as practices that exert power over students in an attempt to control behavior (Guskey, 2004; O'Conner, 2009). These findings may mean that some practices are more likely in classrooms of teachers with low efficacy. A study of 85 high school mathematics teachers examining connections between teachers' level of desired control over students and non-achievement factors in grading (Cicmanec, 1999). Practices that

exert control over students are a product of low teacher efficacy (Woolfolk & Hoy 1990; Bandura 1993). The study however was unable to demonstrate teacher level of control as a predictor of the use of non-achievement factors in grading (Cicmanec, 1999).

Changing long-held grading practices, truly any significant change, is a difficult process. Teachers with higher levels of efficacy are more likely to persist through difficult times and accept greater challenge to make such changes to their practice (Weasmer & Woods, 1998; Cervone, 2000). Findings from a study conducted by Roorda (2008) found professional development with peers resulted in a statistically significant difference the use of achievement and non-achievement factors in grading with greater focus on the use of achievement factors and less on non-achievement factors following professional development. Professional development has also been shown to improve a teacher's level of efficacy (Swars, 2005). Since professional development has had similar impacts on both grading and efficacy, this might mean some relationship between teacher efficacy and grading practices could exist.

While there has been no direct research on grading practices and teacher efficacy, there has been some research connecting teacher's grading practices with the impact on student self-efficacy. Academic self-efficacy in students refers to their perception of their ability to perform on an assessment and it affects the type of assessments students are likely to select, their effort level on those assessments, and how long they are willing to work through difficulties on the assessment (Pintrich, Schunk, 2002). Academic student self-efficacy is impacted by student perception of the assessment tasks. Student perception of assessment tasks is partially determined by the assessment practices of the teacher (McMillian & Workman, 1998).

In a review of literature, Lawrence (2011) notes motivational research frequently indicates the need for specific, accurate feedback from teachers to build student efficacy. In a study of 851 students from 15 high school algebra teachers, researchers found statistically significant results from a single factor ANOVA ($F(4,846) = 2.38$, $p = 2.55$ E-95) indicating that teachers were able to identify different levels of effort from students who earn different letter grades. In the findings, teachers reported the level of effort on a five-point scale for all students resulting in an average score of 4.46 among students earning an A, 3.61 for students who earned a B, 2.91 for students who earned a C, 2.36 for students who earned a D, and 1.42 for students who earned an F. These results show teachers recognizing a decline in effort for students who earned lower grades (Hawkins, 2010). Reporting on student effort may validate students and raise their awareness in turn positively impacting the development of student efficacy (Alkharusi, Aldhafri, Alnabhani, & Alkalbani, 2014).

Inaccurate grades can negatively impact a student's sense of efficacy (Pintrich & Schunk, 2002). Justice perceptions and low performance have been shown to predict low self-efficacy in college students. Justice perceptions are a student's perception of the fairness of the gap between actual and expected performance on an assessment and are derived from the assessment and feedback practices of the teacher (Nesbit & Burton, 2006).

Summary

Grading practices have evolved over the last century through changes at the higher education level that have made their way down to the public school setting. While grades support a number of purposes, the most generally accepted is the use of grades to

report learning progress. In the past few decades, research questioning the validity and reliability of many common grading practices has emerged. Grading practices such as the inclusion of non-achievement factors, the use of zeros on 100-point scales, accepting extra credit, and penalizing scores for missed deadlines are questioned for distorting the final achievement level. Guskey (2004) notes that no research supports the use of negative grades to influence behavior despite many practices intended to just do that.

Teacher efficacy has also been developed as a concept in the past few decades. Stemming from the original work of Bandura in 1977, teacher efficacy refers to a teacher's belief in their ability to influence student achievement. Efficacy is built through personal mastery experiences, the vicarious experiences of respected peers, the social persuasion of colleagues providing encouragement or affirmation, and psychological factors. Teacher efficacy has been linked to student achievement and many teacher dispositions including persistence with difficult students, ownership over achievement results, and controlling practices. There is an absence of research that directly connects a teachers specific grading practices and the level of teacher efficacy.

CHAPTER THREE

METHODOLOGY

This study sought to determine if a relationship existed between a teacher's grading practices and teacher efficacy. The researcher used a survey to collect data from kindergarten through twelfth grade teachers in the southwest region of Missouri. The survey was a compilation of two separate instruments including a grading practices inventory developed by the researcher and an existing instrument for measuring levels of teacher efficacy. Survey results were analyzed to identify if a relationship existed between the two. Results were disaggregated and compared by experience level, grade level taught, size of district, level of education, reporting format and gender.

Research Questions

This study sought to explore if a relationship between teacher efficacy levels and grading practices existed. Further, the study examined trends regarding the levels of effective grading practices and teacher efficacy across various demographics. The following research question guided the study: Is there a relationship between a teacher's grading practices and the level of teacher efficacy?

Subset Questions

1. What are the grading practices of a teacher?
2. What are the efficacy levels of teachers?
3. How do the efficacy levels and grading practices of teachers differ across gender, grade-level taught, district size, years of experience, and level of education?

Participants

Teachers from schools in the southwest region of Missouri were included in this study. Southwest Missouri as defined by the Missouri Department of Elementary and Secondary Education includes 20 counties. These counties contain 388 schools in 94 districts. Within the same are a mixture of school configurations including elementary buildings serving kindergarten through grade six, preschool centers serving only pre-kindergarten, primary centers serving pre-kindergarten through grade two, intermediate schools serving grades three through six, middle schools or junior highs serving grades five through eight, and high schools or technical schools serving grades nine through twelve. Additionally, there are a few small districts with single schools that span kindergarten through grade eight or kindergarten through grade twelve in the same building.

The 94 districts in this region serve 146,941 students. District sizes range from districts as small as 43 students to as large as over 25,000 students. The southwest region mirrors the rest of the state in terms of the number of districts of various sizes. In the southwest region, 40 percent of districts have less than 500 students, 22 percent of district have between 500 and 1,000 students, 33 percent of districts have between 1,000 and 5,000 students, and only 4 percent of districts are larger than 5000 students. In comparison for the entire state of Missouri, 46 percent of districts have fewer than 500 students, 22 percent of district have between 500 and 1,000 students, 25 percent of districts have between 1,000 and 5,000 students, and only 7 percent of districts are larger than 5000 students.

The identified region includes 13,535 certified staff. Teachers included in this study had been teaching from 1 to 20+ years and their level of education ranged from a bachelor's degree to doctorate. Teachers were included who teach any grade or combination of grades from kindergarten through twelfth grade. A demographic question about the participant's responsibility for issuing grades was used to eliminate respondents who were not directly responsible for classroom teaching including instructional coaches, administrators, or other support positions.

Research Design and Process

In this quantitative study, the researcher used two instruments in a single survey sent electronically for distribution to teachers within the target area. The process began by obtaining approval to conduct research from the Research Review Board of Southwest Baptist University. Permission to use the Teacher's Sense of Efficacy Survey was obtained through email consent from Dr. Woolfolk-Hoy and is included in Appendix A. The researcher then began a pilot phase for the Grading Practices Inventory, an instrument developed by the researcher to identify which grading practices a teacher uses. Upon completion of the pilot phase, the final survey was developed electronically through a third-party client and distributed through email. The final survey included an informed consent, demographic questions, the Grading Practices Inventory, and the Teacher's Sense of Efficacy Scale (Short Form). Once data collection ended, results were analyzed.

Pilot Process

Because no instrument had yet been developed to identify specific grading practices among teachers, the researcher developed the Grading Practices Inventory. To ensure the reliability and validity of the instrument, the researcher conducted a pilot test.

Content validity was determined using the index of item objective congruence developed by Rovinelli and Hambleton (1977). An expert panel of six educators with experience in assessment were asked to review the instrument and provide feedback on question clarity and alignment to intended purpose. The panel included classroom teachers with demonstrated teacher-leadership in the area of assessment and grading practices, administrators with backgrounds in assessment, and university professors with experience in instrument design and validity.

Next, the revised instrument was piloted at selected school sites representing both elementary and secondary teachers. Two middle schools and two elementary schools in the Joplin School District were selected and the survey was sent to the principals of those buildings with a request to distribute to their teachers. The initial survey request was made with a reminder follow-up email sent to principals two days later. In all, 32 participants responded to the pilot. Results from the pilot group were analyzed for reliability using Cronbach's Alpha. The 13 items produced an alpha value of .639. To demonstrate construct validity, an exploratory factor analysis was conducted on the pilot results.

Survey Process

Once all instruments were developed and tested, a single survey was created combining both instruments and several questions regarding demographic information.

The survey was created electronically using QuestionPro as a third-party client. The data entry was completed by the individual participants. All responses were stored electronically through the QuestionPro site.

The researcher drafted a cover letter explaining the purpose of the research and use of survey. The letter, including a link to the survey, was emailed to the principals of each school within the target area of southwest Missouri. Part of the letter requested that administrators forward the survey link on to the teachers in their building. The Missouri Department of Elementary and Secondary Education maintains a directory of all public schools, their administrators, and administrator's email addresses making this information easy to obtain. Specific lists of teachers and teacher email addresses are not readily available and for such a large sample size would be difficult to obtain. A follow-up email reminder was sent one week later. In all, 630 complete responses were collected during the administration period.

Research and Ethical Protocols

In accordance with the guidelines of Southwest Baptist University regarding the protection of human participants, a request for review was submitted to the Research Review Board for approval to survey teachers in southwest Missouri. Approval from the Research Review Board was granted.

Potential risks associated with participation in this study are few if any. No personally identifiable information was collected from respondents that would make tracing an individual or specific district possible. All data was collected and maintained through QuestionPro, a third-party client. The survey instrument included an informed

consent as part of the participation. Following collection, data was analyzed and used to make conclusions and recommendations.

Treatment of Data

Data obtained from the Grading Practices Inventory was used to examine the types of practices in place among different teachers. The unweighted mean of all items on the Grading Practices Inventory was used to produce a single overall grading practices score. Likewise, the unweighted mean of all items from the Teacher's Sense of Efficacy Scale provides a single overall efficacy score. Using the Pearson product-moment correlation coefficient, overall efficacy scores as well as scores were compared to the over all grading practices score to determine if a correlation existed. Additionally, an ANOVA was used to determine trends among various subcategories with the demographic information collected.

Instrumentation

To collect data for this study, a single survey comprised of two instruments and a selection of questions regarding demographics was used. The first instrument, the Grading Practices Inventory, was used to identify which specific grading practices a teacher used. The second instrument, the Teacher's Sense of Self-Efficacy Scale, was used to determine the level of efficacy for the teacher. Demographic information collected included grade level primarily taught, size of school district, gender, years teaching experience, and highest degree obtained. The final survey included 15 items on the Grading Practices Inventory, 12 items from the Teacher's Sense of Efficacy Survey, and six demographic questions taking an estimated 15 minutes for an individual participant to complete.

Grading Practices Inventory

The grading practices inventory, a unidimensional instrument, was developed for this study as a way to measure teacher use of specific grading practices identified by grading research. Items for the instrument were developed by reviewing the literature and identifying specific grading practices commonly discussed. Among those outlined in the review of literature provided in Chapter 2 were practices such as norm versus criterion referenced grading; assigning points or percentages versus rubric or proficiency scores; including effort, extra credit, or homework, and adjusting grades based on task difficulty or overall classroom performance. A complete list of items developed is shown in Appendix B. Statements were structured for teachers to rate how frequently they used a particular practice. The structure was modeled after the structure of a similar instrument, the Assessment Practices Inventory developed by Zhang and Burry-Stock (2003).

Because of a desire to create an instrument that would not be overwhelming and cumbersome to the teachers using it, the researcher, with consult from his advisor, consolidated like items to better reflect specific trends in practice. In this way, separate items such as the inclusion of teacher-made tests, textbook assessments, homework, and teacher observations were combined into a single item: "I include evidence of academic performance in student grades (journals, tests, homework, teacher observation)". In all, 36 specific practices were consolidated into 13 items still reflective of the major practices identified in the literature review. The scale was also adjusted to mirror the 9-point scale or frequency used in the Teacher's Sense of Efficacy Survey in an effort to make the final tool less confusing by utilizing a common scale for both instruments. Two additional

items were included to filter responses based on the type of final grade reported and how the teacher's practices were developed.

Validity and reliability of the instrument was determined through the feedback of an expert panel using the index of item objective congruence developed by Rovinelli and Hambleton (1977) and a pilot test. To ensure content validity, an expert panel of six educators with experience in assessment were asked to review the instrument and provide feedback on question clarity and alignment to intended purpose. The panel included two classroom teachers with demonstrated teacher-leadership in the area of assessment and grading practices and four administrators with backgrounds in assessment. Additionally, three university professors with experience in instrument design and validity were consulted. The expert panel was provided the list of 13 items as well as a written description of the idea or practice being measured. Members of the expert panel were then asked to rate each item at +1 for items that closely matched the idea described, 0 for items that do not quite match or are unclear, and -1 for items that do not match the idea described. Experts were further invited to make open comment about wording that may have been confusing. Scores from across all experts were averaged for each item with the intent to revise any item scoring an average agreement rating of 0.8 or less. All items scored above the threshold and were therefore included. Only two minor changes to wording were made as a result of the feedback from the expert panel. In item 2, the word "non-academic" was substituted for "non-achievement" to clarify the focus on academic learning. In item 9, the word "assignments" was added to clarify what grades were adjusted for being late. The final survey submitted for pilot testing is included in Appendix C.

The revised instrument was piloted at selected school sites representing both elementary and secondary teachers and received 32 responses. Results from the pilot group were analyzed for reliability using Cronbach's Alpha. The 13 items produced an alpha value of .639. To demonstrate construct validity, an exploratory factor analysis was conducted on the pilot results. Results of a principal components factor analysis with varimax rotation demonstrated construct validity and supported the uni-dimentionality of the inventory. Those results are shown in Table 1. The final draft of the survey is included in Appendix D.

Table 1: Exploratory Factor Analysis for Grading Practices Inventory

Statements	Component 1
I include evidence of academic performance in student grades (journals, tests, homework, teacher observation).	.593
I include evidence of non-academic performance in student grades (attendance, behavior, participation, effort, attitude).	.223
I assign grades based on student performance against a standard or objective.	.490
I assign grades based on student performance against their own past performance (growth or improvement grading).	.354
I assign grades based on student performance against overall class performance (curving grades).	.503
I determine a final grade by averaging all scores for a given subject, standard or course.	.431
I determine a final grade by considering the median, mode, or other statistical measure for all scores for a given subject, standard or course.	.461
I determine a final grade by using my own professional judgment.	.219
I adjust grades for assignments turned in late.	.647
I penalize a student's grade for not turning in an assignment.	.784
I adjust grades by dropping the lowest or highest score.	.719
I adjust grades for students receiving special education services.	.645
I adjust grades by weighting some scores higher than others.	.604

Note. Principal components with varimax rotation.

After administration of the final version, results were further analyzed to confirm the findings of the initial reliability and validity. To produce a consistent value, items 1, 3, 7, 8 and 9 were recoded. In their initial form, a higher score represented use of

effective grading practices based on research while all others recorded a lower score for most effective practices. After the recode, all items generated a lower score to represent better use of effective grading practices identified in research.

The results of a confirmatory factor analysis failed to load four items with values less than .3. Those results are shown in Table 2. As a result of the results, items 1, 3, 4, and 8 were removed. The results of the remaining items yielded a Cronbach's alpha value of .546.

Table 2: Confirmatory Factor Analysis for Grading Practices Inventory

Statements	Component 1
I include evidence of academic performance in student grades (journals, tests, homework, teacher observation).	.086
I include evidence of non-academic performance in student grades (attendance, behavior, participation, effort, attitude).	.339
I assign grades based on student performance against a standard or objective.	.200
I assign grades based on student performance against their own past performance (growth or improvement grading).	.264
I assign grades based on student performance against overall class performance (curving grades).	.549
I determine a final grade by averaging all scores for a given subject, standard or course.	.388
I determine a final grade by considering the median, mode, or other statistical measure for all scores for a given subject, standard or course.	.350
I determine a final grade by using my own professional judgment.	.263
I adjust grades for assignments turned in late.	.710
I penalize a student's grade for not turning in an assignment.	.644
I adjust grades by dropping the lowest or highest score.	.432
I adjust grades for students receiving special education services.	.471
I adjust grades by weighting some scores higher than others.	.490

Note. Principal components with varimax rotation.

Teacher Sense of Efficacy Scale.

The Teacher Sense of Efficacy Scale (TSES) was developed by Tschannen-Moran, Woolfolk-Hoy, and Hoy (1998). Their scale provides an overall teacher efficacy rating as well as specific ratings across three sub-domains: efficacy in student

achievement, efficacy in instructional strategies, and efficacy in classroom management. The TSES has two forms, a 24-item long form and a 12-item short form. The short form was selected for this study. Permission to use the instrument was obtained through email and confirmed in writing by Dr. Woolfolk-Hoy (personal communication, March 1, 2015) as shown in Appendix A.

To establish the validity of their instrument, the developers of the TSES conducted tests for internal reliability as well as comparisons to other established measures. Results of Cronbach's alpha for the TSES were .94 overall with scores of .87 for the student engagement subscale, .91 for the engagement subscale, and .90 for the classroom management subscale. Externally, Pearson-r for correlation showed the TSES correlated positively to both RAND items (r = .35 and .28) and to the TES developed by Gibson and Dembo (r=.48) (Tschannen-Moran, Woolfolk-Hoy, and Hoy, 1998).

Summary

This quantitative study sought to determine if a connection existed between teacher efficacy and the grading practices of that teacher. Survey data describing teachers' use of grading practices and their level of teacher efficacy was administered to teachers in 20 counties in southwest Missouri. Participants took the already established Teacher's Sense of Efficacy Scale to determine their level of efficacy and the Grading Practices Inventory developed for this study. The resulting data was analyzed to determine trends in grading practices among teacher sub-groups and to identify any correlation between grading practices and a teacher's sense of efficacy.

CHAPTER FOUR

ANALYSIS

Results of the final survey were analyzed to provide insight into the research questions. A total of 630 complete responses were gathered during the survey distribution. This represents a return rate of only 4.6 percent of the population, however, the sample is sufficient to produce a confidence level of 95 percent with a confidence interval of 3.8. Using the unweighted means of items from the Grading Practices Inventory and the Teacher's Sense of Efficacy Scale respectively, a single grading practices and efficacy score was determined and used in further analysis. A Pearson-r was used to determine what correlation exists between the two variables in an attempt to answer the primary research question. Differences between levels of efficacy and use of effective grading practices was examined across the different demographic sub-groups.

Descriptive Statistics

Initially, 737 participants submitted a survey. Of those, 87 were excluded because they answered that they were not responsible for assigning grades and therefore did not qualify to participate in the study. Of the remaining responses, 650 completed the Grading Practices Inventory portion, but only 630 also completed the Teachers Sense of Efficacy Scale. For the purpose of the study, only the 630 who completed both scales in the survey were considered valid. Table 3 shows the means and standard deviations for responses to both scales.

Table 3: Means, Standard Deviations, and Ranges for Scales

Scale	N	Mean	Standard Deviation	Range
Grading Practices Inventory	650	4.70	1.25	1.33 to 7.67
Teacher's Sense of Efficacy Scale	630	7.44	0.90	4.58 to 9.00

Grading Practices Inventory

The mean on the Grading Practices Inventory was 4.70 on a scale that measured from 1 to 9. On the scale, the greater the overall score indicates greater use of ineffective grading practices as determined through current research. The mean score indicates a slight skew toward greater use of ineffective grading practices among participants meaning most of the participants in this study report using less effective grading practices. The standard deviation of 1.25 indicates that the scores are closely grouped near that mean resulting in less variation among the sample. Thus, we can say accurately that the majority of the participants fit this trend and the number of participants at either extreme is low. Reported scores ranged from 1.33 to 7.67 showing a spread of 6.34 across the scale.

Teacher's Sense of Efficacy Scale

The mean for the Teacher's Sense of Efficacy Scale was 7.44 on a scale that measured from 1 to 9. On the scale, the greater the overall score indicates a greater sense of self-efficacy. The mean score indicates most respondents report a high level of efficacy. The standard deviation of .90 indicates even less variation than found in the responses to the Grading Practices Inventory showing that all participants reported efficacy in the mid to high range. Reported scores ranged from 4.58 to 9.00 showing a smaller spread than the Grading Practices Inventory at 4.42. This indicates participants were more consistent in reporting high levels of efficacy across the group than they were at reporting their level of grading practice.

Demographic Data

Beyond the overall comparison of the two scales, results were examined with regard to the various demographic groups selected. Table 4 shows the means and standard deviations for each of the demographic factors included in the survey. The majority of respondents (82 percent) were female. Thirty-six percent of the participants were from districts serving less than 1,000 students and an additional 44 percent were from districts with student populations of 1,000 to 5,000. Only 4 percent were in districts larger than 10,000. The distribution of participants among districts of various size is consistent with the region and state overall. Forty-one percent of participants hold a bachelor's degree and another 54 percent have their master's degree. The sample was evenly distributed between teachers from all levels of experience. Forty-five percent of the participants have been teaching for fewer than 10 years while 21 percent have been teaching for more than 20 years. Participants were spread among grade-levels with 34 percent at the high school level, 40 percent from elementary (either primary or intermediate) and 26 percent from junior high or middle school level (grades 6 to 8).

Table 4: Means and Standard Deviations of Scales by Demographic

		Grading Scale			Efficacy Scale		
Demographic	Factor	N	Mean	SD	N	Mean	SD
Gender	Male	111	5.14	.96	107	7.11	1.01
	Female	539	4.61	1.29	523	7.51	.86
Grade Level	Primary K-2	112	3.85	1.24	107	7.62	.82
	Intermediate 3-5	147	4.25	1.18	140	7.42	.96
	Junior High 6-8	170	4.94	1.20	166	7.47	.96
	High School 9-12	202	5.22	1.02	199	7.35	.85
	Technical 9-12	19	5.42	.78	18	7.23	.86
District Size	1-500	124	4.80	1.21	123	7.34	.89
	500-1,000	110	4.97	1.25	107	7.27	.84
	1,000-5,000	288	4.71	1.20	278	7.51	.91
	5,000-10,000	102	4.27	1.33	96	7.48	.97
	10,000+	26	4.66	1.45	26	7.80	.80
Experience	1-5	163	4.50	1.20	157	7.26	.97
	6-10	130	4.66	1.15	128	7.51	.92
	11-15	126	4.57	1.28	124	7.52	.84
	16-20	93	4.66	1.27	90	7.38	.90
	20+	138	5.11	1.32	131	7.56	.84
Degree	Bachelor's	265	4.64	1.27	259	7.32	.88
	Master's	351	4.74	1.22	340	7.52	.90
	Specialist	29	4.63	1.60	24	7.78	.88
	Doctorate	9	5.10	1.07	7	7.10	1.21

In addition to examining how different demographics performed on the two scales, analysis of the intersection of various demographics also helps to understand the results. Table 5 shows the analysis of how each gender is represented across the different grade spans. There are a greater percentage of female teachers working at the lower grades, while the percentage of male teachers increases at the secondary level. Table 6 shows how the type of grade report varies across different grade levels. A greater percentage of elementary classrooms report using a by-standard method and the percentage of teachers reporting a single overall letter grade increases in the older grades. Given that a elementary schools are more likely to report in a standards-based format, it is likely that differences between males and females found are more likely the result of grade-level taught than gender.

Table 5: Percentage of Genders Across Grade Spans

	Grade Span			
Gender	K-2	3-5	6-8	9-12
Male	1	7	19	29
Female	99	93	81	71

Note. Numbers reported represent percentages.

Table 6: Types of Final Grades Reported by Grade Span

	Grade Span			
Report Type	K-2	3-5	6-8	9-12
Single grade for entire course of subject	17	37	54	59
Separate individual grades for specific competencies	66	38	11	8
Both overall and individual grades for specific competencies	17	35	26	33

Note. Numbers reported represent percentages.

Table 7 shows the means and standard deviations for both scales comparing how teachers developed their current grading practices. Only 13 percent determined their policies completely by personal choice while 38 percent were completely determined by their building or district policies. Thirty-eight percent developed their grading policies by some combination of both personal choice and district or building policy. These results show that a majority of districts in the region appear to be providing some guidance on grading expectations.

Table 7: Means and Standard Deviations of Scales by Origin of Grading Practices

	Grading Scale			Efficacy Scale		
Factor	N	Mean	SD	N	Mean	SD
Personal choice	87	5.18	1.18	83	7.33	.91
Direction of building or district	244	4.39	1.30	234	7.56	.93
Combination of personal choice and direction of building or district	319	4.80	1.18	313	7.39	.87

Table 8 shows the comparison of how policies are developed across districts of various size. The trend shows that in larger districts, a lesser percentage of teachers develop their own policies while a larger percentage are developed by building or district policy. This may be a result of a desire for consistency among teachers in larger districts.

Table 8: Origin of Grading Policies from Districts of Different Size

	District Size In Number of Students				
Origin of Grading Policy	<500	500 to 1,000	1,000 to 5,000	5,000 to 10,000	10,000+
Determined by teacher	23	21	11	2	4
Determine by district	23	30	39	56	46
Determined by combination of teacher and district	54	50	50	42	50

Note. Numbers reported represent percentages.

Inferential Statistics

The analysis of survey data using inferential statistics failed to show significance in the relationship between the two key variables of teacher efficacy and grading practices. However, several significant relationships were discovered among either grading practices or teacher efficacy between different demographic groups. Table 9 shows a summary of the significant findings from the study.

Table 9: Summary of Significant Findings

Test	Factors Compared	Significance
t-test	Gender	$p < .001$
ANOVA	Grading Practices & Grade Taught	$F(4, 645) = 35.77$, $p < .001$
ANOVA	Grading Practices & District Size	$F(4, 645) = 4.53$, $p = .001$
ANOVA	Grading Practices & Experience	$F(4, 645) = 5.18$, $p = .001$
ANOVA	Efficacy & Experience	$F(4, 625) = 2.73$, $p = .029$
ANOVA	Efficacy & Highest Degree	$F(3, 626) = 4.15$, $p = .006$
ANOVA	Grading Practices & Type of Grade Issued	$F(2, 647) = 63.02$, $p < .000$
ANOVA	Grading Practices & Origin of Grading Practices	$F(2, 647) = 15.34$, $p < .000$

Correlation between Grading Practices and Efficacy

The primary research question was to identify is a relationship existed between a teacher's grading practices and his or her level of efficacy. Several possible aspects identified in research suggested that the two factors may share a correlation. To understand the relationship between the grading practices and efficacy levels, the data was first analyzed using the Pearson product-moment correlation coefficient. Individual

participants' mean score from the Grading Practices Inventory was compared to their mean efficacy score from the Teacher's Sense of Efficacy Scale. The results of the Pearson-r indicated no correlation between the two variables, r=.01, n=630, p=.793. This indicates that a teacher's overall grading practices do not significantly correlate with his/her reported level of overall teacher efficacy. Instruments did not allow data to be collected to identify specific grading practices to test if a correlation exists with teacher efficacy.

Relationships in Grading Practices

Further analysis of the variables against demographic groups did result in some statistically significant findings. A t-test shows that the mean difference for grading practices between males and females is significant ($p < .001$) with females (M=4.61) reporting more use of effective practices than males (M=5.14). As indicated earlier, because a majority of females were in the elementary grades while the majority of males were from the secondary grades, this may be more likely the result of the grade level taught than directly related to gender. There was not a significant difference in the means between efficacy levels of the two genders.

Table 10 shows the results of an ANOVA comparing the mean of the grading practices scores across grade-level taught. Results indicated significant differences between grading practices of teachers from Primary K-2, Intermediate 3-5, Middle School/Junior High 6-8, High School 9-12, and Technical School 9-12, $F(4, 645) = 35.77$, $p < .001$. Results of Tukey's HSD post-hoc test showed a significant (p =.042) difference between the grading practices of Primary K-2 (M=3.85) and Intermediate 3-5 (M=4.25) indicating better use of effective grading practices at the lower primary grades.

Additionally, both Primary K-2 and Intermediate 3-5 showed significant ($p < .001$) differences from Middle School/Junior High 6-8 (M=4.94), High School 9-12 (M=5.22), and Technical School 9-12 (M=5.42) indicating both groups of elementary grades demonstrate more effective grading practices than the secondary levels. No significant differences in levels of efficacy reported were found between the grade levels.

Table 10: ANOVA for Grading Practices and Grade-Level Taught

	Sum of Squares	df	Mean Square	F	p value
Between Groups	186.00	4	46.50	35.77	.000*
Within Groups	838.55	645	1.30		
Total	1024.55	649			

Note. *=p<.001

Table 11 shows the results of an ANOVA comparing the mean of the grading practices scores from districts of different size. Results indicated significant differences between grading practices of teachers from districts ranging in size from 1-500 students, 500-1,000 students, 1,000-5,000 students, 5,000-10,000 students and 10,000+ students, $F(4, 645) = 4.53$, $p = .001$. Results of Tukey's HSD post-hoc test showed a significant ($p = .013$) difference between the grading practices of districts with 1-500 students (M=4.80) and 5,000-10,000 students (M=4.27). Additionally, there was a significant ($p < .001$) difference between districts with 500-1,000 students (M=4.97) and 5,000-10,000 students (M=4.27). Further, there was a significant ($p = .021$) difference between districts with 1,000-5,000 students (M=4.71) and districts with 5,000-10,000 students (M=4.27). These results indicate greater use of effective grading practices in larger school districts than in smaller ones. No significant differences in levels of efficacy reported were found between the district sizes.

Table 11: ANOVA for Grading Practices and District Size

	Sum of Squares	df	Mean Square	F	p value
Between Groups	27.91	4	6.99	4.53	.001
Within Groups	996.58	645	1.55		
Total	1024.55	649			

Note. *=p<.001

In Table 12, results of an ANOVA are shown comparing the mean of the grading practices scores from different experience levels of teachers. Results indicated significant differences between grading practices of teachers for teachers who have been teaching 1-5 years, 6-10 years, 11-15 years, 16-20 years, and 20+ years, $F(4, 645) = 5.18$, $p = .001$. Results of Tukey's HSD post-hoc test showed three groups including teachers with 1-5 years experience (M=4.50, $p < .001$), 6-10 years experience (M=4.66, $p = .029$), and 11-15 years experience (M=4.57, $p = .005$) differed significantly from teachers with 20+ years experience (M=5.11) indicating that teachers with more experience tend to use less effective grading practices.

Table 12: ANOVA for Grading Practices and Years Experience

	Sum of Squares	df	Mean Square	F	p value
Between Groups	31.91	4	7.98	5.18	.000*
Within Groups	992.64	645	1.54		
Total	1024.55	649			

Note. *=p<.001

Two questions in the Grading Practices Inventory describe the type of grades teachers issue and how their practices were determined. Participants were asked to describe if they issued a single overall final grade, separate final grades for each standard or competency, or a combination of both. An ANOVA indicated there are significant differences between the grading practices of teachers who report single grades, separate grades, or combination of both methods, $F(2, 647) = 63.02$, $p < .000$. The results of that

ANOVA are shown in Table 13. Tukey's HSD post-hoc test reveals significant ($p <$ .001) differences between the grading practices of teachers who issue separate grades for each standard or competency (M=3.83) and those who issue single overall letter grades (M=5.03) and those who use a combination of both methods (M=4.93) indicating greater use of effective grading practices for teachers who report based on standard or competency only. There was, however, no significant difference between the levels of efficacy for teachers based on the type of final grade they report.

Table 13: ANOVA for Grading Practices and Type of Grade Issued

	Sum of Squares	df	Mean Square	F	p value
Between Groups	167.05	2	83.52	63.02	.000*
Within Groups	857.50	647	1.33		
Total	1024.55	649			

Note. *=p<.001

Table 14 shows the results of an ANOVA comparing grading practices based on how teachers developed their grading policies. The ANOVA results show a significant difference between teachers who developed their policies by personal choice, from a district or building policy, or a combination of both, $F(2, 647) = 15.34$, $p < .000$. Conducting a Tukey's HSD post hoc test showed significant ($p < .001$) differences between teachers developed their policies by personal choice (M=5.18) and by direction of building or district (M=4.39) as well as between those who developed their practices by building or district direction (M=4.39) and a combination of personal choice and building or district direction (M=4.80). Both results indicate that teachers whose policies were developed by building or district direction utilize more effective grading practices than those with a degree of personal choice. There were no significant differences among efficacy levels of teachers based on how their grading practices were developed.

Table 14: ANOVA for Grading Practices and Origin of Grading Practices

	Sum of Squares	df	Mean Square	F	p value
Between Groups	46.39	2	23.19	15.34	.000*
Within Groups	978.17	647	1.51		
Total	1024.55	649			

Note. *=p<.001

Relationships in Efficacy Levels

An ANOVA test for efficacy levels between years experiences showed significant differences among teachers with 1-5, 6-10, 11-15, 16-20 or 20+ years of experience, F(4, 625) = 2.73, p = .029. The results of that ANOVA are shown in Table 15. Only one significant (p = .039) difference showed up in Tukey's HSD post-hoc test between teachers with 1-5 years experience (M=7.26) and those with 20+ years experience (M=7.56) indicating teachers with more experience have greater self-efficacy than those in the first 5 years of their career.

Table 15: ANOVA for Efficacy and Years Experience

	Sum of Squares	df	Mean Square	F	p value
Between Groups	8.77	4	2.19	2.73	.029
Within Groups	503.00	625	.81		
Total	511.77	629			

Note. *=p<.001

No significant differences appeared between the grading practices of teachers based on their level of education. There were, however, significant differences in the efficacy levels of teachers based having a bachelor's, master's, specialist or doctorate degree, F(3, 626) = 4.15, p = .006. Results of Tukey's HSD post-hoc test showed a significant (p = .027) difference between teachers with their bachelor's degree (M=7.32) and master's degree (M=7.52) indicating greater sense of efficacy at the master's level

than the bachelor's. Table 16 shows the results of the ANOVA for efficacy levels and degree earned.

Table 16: ANOVA for Efficacy and Degree Earned

	Sum of Squares	df	Mean Square	F	p value
Between Groups	9.98	3	3.33	4.15	.006
Within Groups	501.79	626	.80		
Total	511.77	629			

Note. *=p<.001

Summary

Analysis of the relationship of overall levels of efficacy and general use of effective grading practices shows no significant correlation between the two factors. Further analysis did revealed statistically significant differences in the use of effective grading practices. Females reported using more effective grading practices than males. Teachers at the elementary level reported using more effective practices than teachers at middle school or high school levels. Teachers in larger school districts reported using more effective grading practices than teachers in smaller districts. Teachers with 20 or more year experience reported using less effective grading practices than teachers with less experience. Grading practices appear to be influenced by the type of reporting format used and the origin of grading practices. Teachers reported greater use of effective practices when grades are reported in a standards-based format and when districts or building have established policies for teachers to follow. Confirming existing research, teachers with greater experience reported greater efficacy than those teachers with less experience. Teachers with master's degrees also reported higher levels of efficacy than those with only a bachelor's degree. Conclusions based on these findings and recommendations for further study are presented in Chapter 5.

CHAPTER FIVE

CONCLUSIONS

Because grading is such an integral part of a teacher's daily work, it seems possible that the practices associated with grading would share a relationship with a teachers' sense of self-efficacy. Results of data analysis indicated no clear relationship between grading practices and efficacy levels. However, analysis of grading practices and efficacy separately across various demographics and reporting formats show some significant patterns. Understanding these patterns provide the basis for recommendations regarding training and policy for grading practices.

Conclusions

The primary research question in this study was to determine if a relationship exists between a teacher's grading practices and the level of teacher efficacy. Results of a Pearson's-r demonstrated that there was no significant correlation between grading practices and levels of efficacy among the participants in this study. This means the researcher was unable to reject the null hypothesis that there is no relationship between a teacher's grading practices and the level of teacher efficacy. Though grading does create a representation of a student's achievement and teacher self-efficacy is a reflection of a teacher's belief in his or her ability to impact student achievement, there does not appear to be any connection to the ways in which teachers grade and the resulting level of efficacy.

Grading Practices

A secondary question in this study was to describe the grading practices of teachers. While the survey instrument itself was only capable of producing a general

grading practices score indicating use of research-based effective grading practices, several trends appeared regarding the likelihood of different demographics to use effective grading practices. Examining these trends can help to understand the landscape of grading reform in southwest Missouri.

Females were more likely to practice effective grading practices than males. However, most males in this study were in the secondary level and the elementary level participants were predominately female. Also, teachers at the kindergarten to fifth grade levels were more likely to report by standard while teachers in grades 6 through 12 were more likely to give a single overall grade. It seems more likely that the increased use of effective grading practices has more to do with grade level taught or reporting format than gender. Elementary grades have been demonstrated in other research to be more likely to report by standard while secondary grades report more often by overall letter grade (Freedman & Frisbie, 1995; Marzano, 2000), so this study is consistent with previous research. The increased use of single overall letters of grades at the upper levels stems from the desire to compare students (Huntsinger & Jose, 2009), determine class rank and honors or athletic eligibility (Marzano, 2000), and for application to post-secondary institutions (Talley & Mohr, 1993).

Many of the grading practices deemed ineffective by research presented would be less likely to be present when teachers must report academic achievement based on individual standards instead of an overall grade. The concern of hodgepodge grading (Guskey, 2006, 2011) by including non-achievement factors in the final grade may be reduced when teachers are attending more directly to how students are progressing against specific criteria for given standards. Averaging all performance factors into a

single final grade lends itself more easily to including such non-achievement factors as attendance, behavior, or effort. In the same way, the effective grading practice of reporting work habits and behaviors as separate grades seems more likely cases where teachers are already reporting academic progress in separate categories by standard.

More use of effective grading practices was also found in districts of larger size. One possible explanation is the amount of professional development and the level of coordination a larger district would have over a smaller district. Research indicates a lack of training of quality measurement and assessment techniques (Stiggins, Frisbie, & Griswold, 1989; Zhang & Burry-Stock, 2003; Wolfe, Viger, Jarvinen & Linksman, 2007), so the ability to provide increased professional development may be why teachers from larger school districts have better practices. Another common factor among those districts of larger size was the increased number of teachers who developed their grading practices based on district or building policy. The expectation of a district-wide policy would provide greater consistency among teachers grading practices but, if informed by quality professional development on research-based assessment practices, would also provide greater direction to teachers who do not have sufficient background in assessment and grading research themselves.

Teachers with 15 years' experience or less showed a more statistically significant likelihood to use effective grading practices than those who had been teaching for 20+ years. While grading reform has been an issue in the past, this may be an indication of better training in teacher preparation programs regarding measurement and assessment techniques. It may also be the result of a greater adaptability to changing demands and

current trends for newer teachers than ones who have been teaching for a longer time and may be more set in their methods.

Teacher Efficacy

Another secondary question in this study was to describe the levels of efficacy for teachers who implement various grading practices. The survey instrument was not specific enough to isolate individual practices for comparison, however, it was able to determine teachers' overall use of effective research practices. For most of the demographic areas examined, there were no statistically significant differences in the levels of efficacy. Only two areas showed significant differences in efficacy levels. In years of experience, teachers with 1-5 years reported lower efficacy that those who had been teaching for 20+ years. This appears consistent with trends that over 40 percent of teachers leave the profession in their first five years (Farber, 1991). It also supports the idea that the mastery experiences accrued over a 20+ year career serve to reinforce higher levels of efficacy than those teachers with less experience. The second area, degree earned, showed a significantly higher level of efficacy for teachers with their master's degrees over those with only a bachelor's. This finding may be the result of additional training providing greater mastery experiences and in turn higher efficacy or may simply occur because teachers who have their master's degree are more likely to have been teaching longer than five years.

Recommendations

The results of this study support claims made in other research (Stiggins, Frisbie, & Griswold, 1989; Zhang & Burry-Stock, 2003; Wolfe, Viger, Jarvinen & Linksman, 2007) of the need for professional development for teachers in the area of grading and

assessment and continued reform in reporting methods. Results indicate a greater use of effective practices in larger districts where professional development opportunities are greater possibly indicating the need for more professional development for smaller districts. Effective grading practices are found more often in environments where grades are reported separately, however, without support and training many teachers may lack the necessary understanding to identify and avoid ineffective grading practices. Teachers may continue to adhere to ineffective practices for the control and behavioral motivational impact they appear to have (Marzano, 2000; Guskey, 2006; Liu, 2007). Before teachers can adjust their grading practice, they would need a deeper understanding of effective grading practices as well as motivational methods for students. There also appears to be value in districts developing research-based policies regarding grading practices. District policy provides consistency for teachers, is shown in this study to be connected to greater use of effective grading practices, and may result in additional professional development for teachers during the development and implementation phases of the policy. This calls for greater education on the part of school leaders to understand measurement and reporting research as they lead grading reform in their schools. On a larger scale, community and societal education regarding the ineffectiveness of what are viewed as traditional grading practices may be necessary to prevent backlash and ease pressure on teachers during their learning phase.

This study also provides opportunities for further research in the areas of grading practices and efficacy. Further testing of the Grading Practices Inventory developed for this study is in order. Several items failed to load during final administration of the survey. Additionally, reliability results were lower for the final test than in pilot testing.

Further studies might also look closer at the relationship of specific grading practices and teacher efficacy. Though no correlation was identified between general grading practices and teacher efficacy, the development of a tool that can be used to identify specific grading practices in comparison to efficacy levels may have clearer results. Results for the difference between levels of efficacy and reporting format were approaching significance in this study of general practices. Another avenue to explore might be to examine efficacy levels of teachers before, during and after implementation of grading reform to see if individual teachers experience a change in efficacy through the process. Understanding what relationships may exist would be beneficial for leaders in preparing to implement grading reform in their buildings or districts.

Summary

Seeking to understand the relationship between grading practices and teacher efficacy could be a valuable tool for educational leaders preparing to implement grading reform. No correlation was discovered between a teacher's general grading practices and their level of efficacy. However, relationships between the use of more effective grading practices and grade level taught, district size, reporting format, or origin of practice may help school leaders in improving the quality of grading practices. Relationships show that lower grades where standards-based formats are more common report greater use of effective practices. It can be inferred that the reporting format influences the types of practices a teacher uses and that selecting a reporting format which provides details of student progress on individual standards requires teachers to adopt more effective grading practices. Another relationship shows larger districts where standardized policies are more common have more teachers using effective grading practices. This suggests that

eliminating the variability inherent among teachers by establishing research-based grading policies can improve the overall grading practices of teachers and calls for the education of school leaders in the effective practices they might then guide teachers to use. Relationships between teacher efficacy and level of education or years experience support existing research about the development of efficacy in teachers. Additionally, opportunities for future research may yet identify a relationship that can better inform school leaders in their reform efforts.

References

Alcock, M. (2003). *A teacher self-efficacy content analysis of fourteen-years of education digest: Are we paying attention?*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (305210933)

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Classroom assessment: Teacher practices, student perceptions, and academic self-efficacy beliefs. *Social Behavior & Personality: An International Journal*, *42*(5), 835-855.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *Clearing House: A Journal o f Educational Strategies, Issues and Ideas. 78*(5). 218-223.

Allen, J. D. & Lambating, J. (2001, April). *Validity and reliability in assessment and grading: Perspectives of preservice and inservice teachers and teacher education professors.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*(2)*,* 86-95.

Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research, 34*(2), 148-165.

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, E., & Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools*. Retrieved from ERIC database. (130 243).

Ashton, P., & Webb, R. (1982, March). *Teacher's sense of efficacy: Toward an ecological model.* Paper presented at the annual meeting of the American Educational Research Association in New York.

Ashton, P., & Webb, R. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York, NY: Longman.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2)*,* 122-147

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Bandura, A. (2000). Cultivate self-efficacy for personal and organizational effectiveness. In E. A. Locke (Ed.), *Handbook of principles of organization behavior.* (pp. 120-136). Oxford, UK: Blackwell.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.) *Self-efficacy beliefs of adolescents*. (pp. 307-337). Charlotte, NC: Information Age Publishing.

Bandura, A. (Ed.) (1995). *Self-efficacy in changing societies.* New York: Cambridge.

Barnes, L. B. (1997). Development of the faculty beliefs about grades inventory. *Educational & Psychological Measurement*, *57*(3), 459-468.

Birnbaum, R. (1977). Factors related to university grade inflation. *Journal of Higher Education*, 5, 519-538.

Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership, 49*(6) 15-18

Blazevski, J. (2006). *Teacher efficacy supporting teacher motivation.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (3224822)

Bonesronning, H. (2004). Do the teachers' grading practices affect student achievement?. *Education Economics, 12*(2)*,* 151-167.

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*(3)*,* 139-154.

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research and Evaluation*, *16*(17), 1–12.

Brookhart, S. M. (1994). Teacher's grading: Practice and theory. *Applied Measurement in Education, 7*(4)*,* 279-301.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort. *Applied Measurement in Education, 10*(2), 161-180.

Brookhart, S. M. (2008). *How to give effective feedback to your students.* Alexandria, VA: Association for the Supervision and Curriculum Development.

Brookhart, S. M. (2011). *Grading and learning: Practice that support student achievement*. Bloomington, IN: Solution Tree Press.

Brown, C. G. (2012). A systematic review of the relationship between self-efficacy and burnout in teachers. *Educational and Child Psychology, 29*(4), 47-63.

Brown, I., & Inouye, D. K. (1978). Learned helplessness through modeling: The role of perceived competency in competence. *Journal of Personality and Social Psychology, 36,* 900-908.

Burke, R.J. (1968). Student reactions to course grade. *The Journal of Experimental Education* 36 (4), 11-13.

Bursuck, W. D., Munk, D. D., & Olson, M. M. (1999). The fairness of report card grading adaptations: What do students with and without learning disabilities think?. *Remedial & Special Education*, *20*(2), 84-92.

Bursuck, W. D., Polloway, E. A., Plante, L. L., Epstein, M. H., Jayanthi, M. M., & McConeghy, J. J. (1996). Report card grading and adaptations: A national survey of classroom practices. *Exceptional Children*, *62*(4), 301-318.

Calik, T., Sezgin, F., Kavgaci, H., & Kilinc, A. C. (2012). Examination of relationships between instructional leadership of school principals and self-efficacy of teachers and collective teacher efficacy. *Educational Sciences: Theory & Practice*, *12*(4), 2498-2504.

Campbell, A.L. (1921). Keeping the score. *The School Review* 29 (7), 510-520.

Canady, R. L., & Hotchkiss, P. R. (1989). It's a good score! Just a bad grade. *Phi Delta Kappan, 71*(1), 68-71.

Cervone, D. (2000). Thinking about self-efficacy. *Behavior Modification, 24,* 30-57.

Cicmanec, K. B. M. (1999). *High school mathematics teachers: Grading practice and pupil control ideology.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (304515742).

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2)*,* 159-179.

Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education, 60*(4), 323-337.

Collins, C. (2006). *A mixed method investigation of promoting self-efficacy in a teacher induction program.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (305355189)

Cox, K. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67-87.

Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self. *Teaching & Teacher Education, 24*(3), 751-766.

Downes, T. (1993). Student-teachers' experiences in using computers during teaching practice. *Journal of Computer Assisted Learning*, *9*(1), 17-33.

Dueck, M. (2014). *Grading smarter not harder: Assessment strategies that motivate kids and help them learn.* Alexandria, VA: Association for the Supervision and Curriculum Development.

Enderlin-Lampe, S. (2002) Empowerment: Teacher perceptions, aspirations, and efficacy. *Journal of Instructional Psychology, 29*(3), 139-146.

Evers, W. J. G; Brouwers, A. & Tomie, W. (2002). Bumout and self-efficacy: A study on teachers' beliefs when implementing an innovative educational system in the Netherlands. *British Journal of Educational Psychology, 72*(2)*,* 227-243.

Farber, B. A. (1991). *Crisis in education: Stress and burnout in the American teacher.* Jossey-Bass. Abstract retrieved from http://psycnet.apa.org.

Friedman, S. J., & Frisbee, D. A. (1993, April). *The validity of grade cards as indicators of student performance.* Paper presented at the meeting of the National Council on Measurement in Education, Atlanta, GA.

Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational & Psychological Measurement*, *55*(1), 5-26.

Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practices, 11*(3), 35-42.

Froman, R.D. & Owen, S.V. (1989) Infant care self-efficacy. *Scholarly Inquiry for Nursing Practice: an International Journal 3*(3), 199–210.

Geyer, G.H. (1938). Westwood high goes off the 'gold-standard' marking system. *The Clearing House* 12 (9), 530-533

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569–582.

Gillett, T., & Urbanski, A. (2004). The union's role in teachers' careers. In J. I. Goodlad & T. J. McMannon (Eds.), *The Teaching Career* (pp. 140-56). New York, NY: Teachers College Press.

Glidewell, J.C. & Livert, D.E. (1992) Confidence in the practice of clinical psychology. *Professional Psychology: Research and Practice 23*(5), 362–368.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37* (2), 479-507.

Goddard, R., Hoy, W., & Woolfolk Hoy, A. (2004). Collective efficacy: Theoretical development, empirical evidence, and future directions. *Educational Researchers, 33*(3), 2-13.

Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching*, *58*(3), 93-98.

Grimes, T. V. (2010). Interpreting the meaning of grades: A descriptive analysis of middle school teachers' assessment and grading practices. (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses, (305268025).

Guskey, T. R. (1980). Mastery learning: Applying the theory. *Theory Into Practice*, *19*(2), 104-111.

Guskey, T. R. (1994). Making the grade: What benefits students?. *Educational Leadership*, *52*(2), 14-20.

Guskey, T. R. (2001a). Fixing grading policies that undermine standards. *Education Digest*, *66*(7), 16-21.

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

Guskey, T. R. (2001c). High percentages are not the same as high standards. *Phi Delta Kappan*, *82*(7), 534-536.

Guskey, T. R. (2002). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan*, *83*(10), 775-780.

Guskey, T. R. (2004). Are zeros your ultimate weapon?. *Education Digest*, *70*(3), 31-35.

Guskey, T. R. (2005). Mapping the road to proficiency. *Educational Leadership*, *63*(3), 32-38.

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, *87*(9), 670-675.

Guskey, T. R. (2007a). Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for mastery". *Journal Of Advanced Academics*, *19*(1), 8-31.

Guskey, T. R. (2007b). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues & Practice*, *26*(1), 19-27.

Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, *69*(3), 16-21.

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

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press.

Guskey, T. R., & Jung, L. (2009). Grading and reporting in a standards-based environment: Implications for students with special needs. *Theory Into Practice*, *48*(1), 53-62.

Harrison, M. A., Meister, D. G., & Lefevre, A. J. (2011). Which students complete extra-credit work?. *College Student Journal*, *45*(3), 550-555.

Hawkins, J. P. (2010). *The impact of teacher perception of student effort and student attitude on grading of secondary students.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (816908566).

Healy, C. and S. Rojstaczer. (2012). Where a is ordinary: The evolution of American college and university grading: 1940-2009. *Teachers College Record* 114 (7), 1-23.

Henderson-Montero, D., Julian, M. W., & Yen, W. M. (2003). Multiple perspectives on multiple measures: An introduction. *Educational Measurement: Issues and Practice, 22*(2), 7–12.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: a longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774-786.

Hoy, A. (2000) *Changes in teacher efficacy during the early years of teaching.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Hoy, A., & Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*, 343-35.

Hoy, W., & Sabo, D. (1998). *Quality middle schools: Open and healthy*. Thousand Oaks, CA: Corwin Press.

Hughes, J. N., Barker, D., Kemenoff, S., & Hart, M. (1993). Problem ownership, causal attributions, and self-efficacy as predictors of teachers' referral decisions. *Journal of Educational Psychological Consultation, 4*(4)*,* 369-384.

Huntsinger, C. S. & Jose, P. E. (2009). Parental involvement in children's schooling: Different meanings in different cultures. *Early Childhood Education Quarterly 24*(4), 398-410.

Jex, S. M. (1999). Efficacy beliefs as a moderator of the impact of work-related stressors: A multilevel study. *Journal of Applied Psychology, 84*(5)*,* 349-361.

Jhanke, M. S. (2010). *How teacher collective efficacy is developed and sustained in high*

achieving middle schools. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Thesis. (3410320)

Kaestle, C. (1983). *Pillars of the republic: Common schools and American society, 1780-1860.* New York: Hill and Wang.

Kliebard, H.M. (1987). *The struggle for the American curriculum, 1893-1958*. New York: Routledge.

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise and other bribes*. Boston, MA: Houghton Mifflin.

Kohn, A. (1999). From grading to degrading. *High School Magazine 6*(5), 38-48.

Kvaraceaus, W.C. (1939). The primary report card. *The Elementary School Journal* 39 (10), 747-750.

Lambating, J. & Allen, J. D. (2002, April). *How the multiple functions of grades influence their validity and value as measures of academic achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Lawrence, J. A. (2011). *An examination of high school assessment and grading practices within a professional learning community.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (874246041).

Liu, X. (2007). *Measuring teachers' perceptions of grading practices: A cross-cultural perspective.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (304862719).

MacIver, D. (1990). *A national description of report card entries in the middle grades.* Retrieved from ERIC database. (ED324124).

Maginnis, J. L. (2009). *The relationship clinical faculty training has to student teacher self-efficacy.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (305092871)

Marachi, R., Gheen, M. Midgley, C. (2001). An examination of elementary, middle, and high school teachers' beliefs and approaches to instruction using a goal theory framework. Unpublished Manuscript, University of Michigan, Ann Arbor.

Margolis, H., & McCabe, P. P. (2006). Improving self-efficacy and motivation: What to do, what to say. Intervention in School & Clinic, 41(4), 218-227.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

McMillan, J. H., & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Retrieved from ERIC database. (ED453263).

Moore, W., & Esselman, M. (1992, April). *Teacher efficacy, power, school climate and achievement: A desegregating district's experience*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Morris, V. (1995). *Parent efficacy, teacher efficacy, and parent involvement in professional development schools.* Research Report for Frayser Elementary School. Center for Research in Educational Policy, The University of Memphis, College of Education, Memphis, TN 38152.

Munk, D., & Bursuck, W. D. (1998). Report card grading adaptations for students with disabilities: Types and acceptability. *Intervention in School and Clinic, 33,* 306-308.

National Education Association. (1971). Reporting pupil progress to parents. *National Education Association Research Bulletin* 49, 81-83.

Neck, C. P., & Manz, C. C. (1992). Thought self-leadership: the influence of self-talk and mental imagery on performance. *Journal of Organizational Behavior*, *13*(7), 681-699.

Nesbit, P. L., & Burton, S. (2006). Student justice perceptions following assignment feedback. *Assessment & Evaluation In Higher Education*, *31*(6), 655-670.

Newmann, F., Rutter, R., & Smith, M. (1989). Organizational factors that affect school sense of efficacy, community, and expectations. *Sociology of Education, 62*(4), 221-238.

O'Conner, K. (2009). *How to grade for learning*. Thousand Oaks, CA: Corwin.

O'Conner, K. (2011). *A repair kit for grading.* Boston, MA: Pearson Education.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4)*,* 543-578.

Peabody, A. (1888). *Harvard reminiscences.* Boston: Ticknor.

Percell, J. C. (2014). *Essentially point-less: The influence of alternative, non points-based grading on teachers' instructional practices.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (1622414792).

Pintrich, P., & Schunk, D. (1996). *Motivation in education* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Polloway, E. A., Epstein, M. H., Bursuck, W. D., Roderique, T. W., McConeghy, J. J., & Jayanthi, M. M. (1994). Classroom grading; A national survey of policies. *Remedial And Special Education*, 15(3), 162-170.

Putman, M. S. (2012). Investigating teacher efficacy: Comparing preservice and inservice teachers with different levels of experience. *Action in Teacher Education, 34*(1), 27-40.

Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *Journal Of Educational Research*, *102*(3), 175-186.

Reece, S. M., & Harkless, G. (1998). Self-efficacy, stress, and parental adaptation: Applications to the care of childbearing families. *Journal of Family Nursing, 4*(2), 198-215.

Reece, S.M. (1993) Social support and the early maternal experience of primiparas over 55. *Maternal Child Nursing Journal 21*(3), 91–98.

Ring, M. M., & Reetz, L. (2000). Modification effects on attribution of middle school students with learning disabilities. *Learning Disabilities Research & Practice, 15*, 34–42.

Roorda, N. L. (2008). *The impact of professional development on assessment and grading practices for secondary teachers.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (288264353).

Ross, J. A. (1992). Teacher efficacy and the effect of coaching on student achievement. *Canadian Journal of Education, 17(1)*, 51-65.

Ross, J. A. (1994). *Beliefs that make a difference: The origins and impacts of teacher*

efficacy. Paper presented at the Annual Meeting for Curriculum Studies, Calgary

Ross, J. A. (1995). Strategies for enhancing teachers' beliefs in their effectiveness: Research on a school improvement hypothesis. *Teachers College Record, 97*(2), 227-252.

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 2*, 49-60.

Rundquist, A. (2012). Standards-based grading with voice: Listening for students' understanding. *AIP Conference Proceedings*, *1413*(1), 69-72.

Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A-F marking scheme. *Journal Of Curriculum Studies*, *46*(2), 201-224.

Schumacher, D. F. (2009). *Collective teacher efficacy and student achievement*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (3355838)

Schwartz, K. A. (2010). *Dynamics of teacher self-efficacy: Middle school reading and language arts teacher responses on a teacher sense of efficacy scale* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (822778362)

Selby, D. & Murphy, S. (1992). Graded or degraded: Perceptions of letter grading for mainstreamed learning-disabled students. *British Columbia Journal of Special Education, 16*(1), 92-104.

Sibbald, T. M. (2008). *The connection between teacher self-efficacy and reflective practice.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (304361689)

Siegle, D., & McCoach, D. B. (2007). Increasing student mathematics self-efficacy through teacher training. *Journal of Advanced Academics, 18*(2), 278-312.

Skaalvik, E. M. & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611-625

Snyder, T.D. (ed.) (1993). *120 years of American education: A statistical portrait* Washington, DC: National Center for Education Statistics.

Spady, W. G. (1991). Shifting the grading paradigm that pervades education. *Outcomes*, *6*(1), 7-12.

Sparks, G. (1988). Teacher attitudes toward change and subsequent improvement in classroom teaching. *Journal of Educational Psychology, 80*(1), 111-117.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high school work in English, *School Review, 21,* 254-295.

Starch, D., & Elliott, E. C. (1913). Reliability of grade work in mathematics. *School Review, 21,* 676-681.

Stiggins, R. J. (1992), High quality classroom assessment: What does it really mean? *Educational Measurement: Issues and Practice, 11*(2), 35-39.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14.

Swars, S. L. (2005). Examining perceptions of mathematics teaching effectiveness among elementary preservice teachers with differing levels of mathematics teacher efficacy. *Journal of Instructional Psychology, 32*(2), 139-147.

Talley, N. R. & Mohr, J. I. (1993). The case for national standard of grade weighting. *Journal of College Admissions, 139,* 9-13.

Tocci, C. (2010). An Immanent Machine: Reconsidering grades, historical and present. *Educational Philosophy & Theory*, *42*(7), 762-778.

Tschannen-Moran, M. & Barr, M. (2004). Fostering student learning: The relationship of collective teacher efficacy and student achievement. *Leadership and Policy in Schools, 3*(3), 189-209.

Tschannen-Moran, M. & McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *The Elementary School Journal, 110*(2), 228-245.

Tschannen-Moran, M. & Woolfolk Hoy, A. (2006). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*(6), 944-956.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive concept. *Teaching and Teacher Education, 17*(7)*,* 783-805.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2)*,* 202-248.

Tuchman, S. E. (2010). *The influence of formative pre-service experiences on the teacher self-efficacy of Jewish day school teachers.* . (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (305234554)

Tucker, C. M., Porter, T., Reinke, W., Herman, K. C., Ivery, P., Mack, C., & Jack, E. (2005). Promoting teacher efficacy for working with culturally diverse students.

Preventing School Failure, 50(2), 29-34.

Valdes, K. A., Williamson, C. L., & Wagner, M. M. (1990). *The national longitudinal transition study of special education students.* Retrieved from ERIC database. (ED324893).

Viaderi, D. (2005). Researchers connect lower achievement, high school exit tests. *Education Week, 24*(21), 10.

Waltman, K. K., & Frisbie, D. A. (1994, April). *Absolute and relative grading standards*. Paper presented at the meeting of the American Educational Research Association, Atlanta, GA.

Wang, L., Ertmer, P. A., & Newby, T. J. (2004). Increasing preservice teachers' self-efficacy beliefs for technology integration. *Journal of Research on Technology in Education, 36*(3), 231-250.

Watson, S. (1991). *A study of the effects of teacher efficacy on academic achievement of third- grade students in selected elementary schools in South* Carolina. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (9230552)

Weasmer, J. & Woods, A. M. (1998). I think I can: The role of personal teaching efficacy in bringing about change. *Clearing House, 71*(4)*,* 245-247.

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 & Practice*, *32*(2), 26-36.

Wiggins, G., & McTighe, J. (2006). *Understanding by design*. Upper Saddle River, NJ: Pearson Education.

Wolfe, E. W., Viger, S. G., Jarvinen, D. W., & Linksman, J. (2007). Validation of scores from a measure of teachers' efficacy toward standards-aligned classroom assessment. *Educational & Psychological Measurement*, *67*(3), 460-474.

Woolfolk, A. E. (2001). *Educational psychology. (7thEdition).* Needham Heights, NY: Allyn & Bacon.

Woolfolk, A. E. & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*(1)*,* 81-91.

Woolfolk, A., Rosoff, B., & Hoy, W. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching & Teacher Education, 6,* 137-148.

Wright, D., & Wise, M. J. (1988). Teacher judgment in student evaluation: A comparison of grading methods. *Journal of Educational Research, 82*(1), 10-14.

Yeo, L., Ang, R., Chong, W., Huan, V., & Quek, C. (2008). Teacher efficacy in the Context of Teaching Low Achieving Students. *Current Psychology, 27*(3), 192-204.

Zeldin, A. L. & Pajares, F. (1997, March). *Against the odds: Self-efficacy beliefs of women with math-related careers.* Paper presented at the meeting of the American Educational Research Association, Chicago.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement In Education*, *16*(4), 323-342.

Appendix A: Permission for Use of Teacher Sense of Efficacy Scale

THE OHIO STATE UNIVERSITY COLLEGE OF EDUCATION AND HUMAN ECOLOGY

ANITA WOOLFOLK HOY, PH.D.

PROFESSOR
PSYCHOLOGICAL STUDIES IN EDUCATION

Dear Jason Weaver:

You have my permission to use the *Teachers' Sense of Efficacy Scale* in your research. A copy the scoring instructions can be found at:

http://u.osu.edu/hoy.17/research/instruments/

Best wishes in your work,

Anita Woolfolk Hoy

Anita Woolfolk Hoy, Ph.D.
Professor Emeritus

COLLEGE OF EDUCATION
29 WEST WOODRUFF AVENUE
COLUMBUS, OHIO 43210-1177

WWW.COE.OHIO-STATE.EDU/AHOY

PHONE 614-292-3774
FAX 614-292-7900
HOY.17@OSU.EDU

Appendix B: First Draft of Grading Practices Inventory

The Grading Practices Inventory is designed to gather data about the specific grading practices of individual teachers. To respond to items, consider your own classroom grading and the final grades you issue to students.

Which best describes your classroom?

- ☐ general education classroom with regular education students only
- ☐ general education classroom with both regular and special education students
- ☐ special education classroom with only special education students

What type of final grades do you provide for each student?

- ☐ a single grade for the entire course or subject area
- ☐ separate individual grades for specific competencies, skills or standards in the course or subject
- ☐ both an individual grades for specific competencies, skills or standards as well as a single overall summary grade for the course or subject area

How were your grading policies and practices determined?

- ☐ by personal choice
- ☐ by direction of a building or district grading policy
- ☐ by a combination of building or district grading policy and personal choice

For each of the following statements, rate how frequently you do the following: Never, Seldom, Occasionally, or Often.

- I reduce a grade for being turned in late.
- I record a zero for students who do not turn in assigned work.
- I record scores as a total number of points earned.
- I record scores as a percentage of correct responses.
- I record scores based on a 3, 4, or 5 point rubric or other similar limited scale.
- I assign grades based on student performance against a standard or objective.
- I assign grades based on student performance against their own past performance (improvement grading).
- I assign grades based on student performance against overall class performance (curving grades).
- I adjust grades for students receiving special education services.
- I weight various types of assignments or tests differently.
- I drop the lowest score from a set.
- I drop the highest score from a set.
- I determine a final grade by averaging all scores for a given subject, standard or course.
- I determine a final grade by considering the median, mode, or other statistical measure for all scores for a given subject, standard or course.
- I determine a final grade using professional judgment.
- I adjust grades based on difficulty of the task.

- I adjust grades based on student effort.
- I adjust grades based on overall class performance.
- I adjust grades based on amount of improvement an individual student has made.
- I include actual academic performance in determining final grades.
- I include attendance in determining final grades.
- I include student behavior in determining final grades.
- I include group grades in determining final grades.
- I include participation in determining final grades.
- When assigning a final grade, I include extra credit.
- When assigning a final grade, I include homework.
- When assigning a final grade, I include individual projects.
- When assigning a final grade, I include group projects.
- When assigning a final grade, I include teacher-made tests.
- When assigning a final grade, I include tests written by textbook companies.
- When assigning a final grade, I include student journals.
- When assigning a final grade, I include student notebooks.
- When assigning a final grade, I include teacher observation.
- When assigning a final grade, I include student self-reflection.
- When assigning a final grade, I include student self-assigned grades.
- When assigning a final grade, I include student peer-assigned grades.

Appendix C: Second Draft of Grading Practices Inventory (Expert Pilot)

Grading Practices Inventory

The Grading Practices Inventory is designed to gather data about the specific grading practices of individual teachers. To respond to items, consider your own classroom grading and the final grades you issue to students.

What type of final grades do you provide for each student?

- ☐ a single grade for the entire course or subject area
- ☐ separate individual grades for specific competencies, skills or standards in the course or subject
- ☐ both an individual grades for specific competencies, skills or standards as well as a single overall summary grade for the course or subject area

How were your grading policies and practices determined?

- ☐ by personal choice
- ☐ by direction of a building or district grading policy
- ☐ by a combination of building or district grading policy and personal choice

	Grading Practice	Never		Seldom		Sometimes		Quite a bit		A Great Deal
1	I include evidence of academic performance in student grades (journals, tests, homework, teacher observation).	1	2	3	4	5	6	7	8	9
2	I include evidence of non-academic performance in student grades (attendance, behavior, participation, effort, attitude).	1	2	3	4	5	6	7	8	9
3	I assign grades based on student performance against a standard or objective.	1	2	3	4	5	6	7	8	9
4	I assign grades based on student performance against their own past performance (growth or improvement grading).	1	2	3	4	5	6	7	8	9
5	I assign grades based on student performance against overall class performance (curving grades).	1	2	3	4	5	6	7	8	9
6	I determine a final grade by averaging all scores for a given subject, standard or course.	1	2	3	4	5	6	7	8	9
7	I determine a final grade by considering the median, mode, or other statistical measure for all scores for a given subject, standard or course.	1	2	3	4	5	6	7	8	9
8	I determine a final grade by using my own professional judgment.	1	2	3	4	5	6	7	8	9
9	I adjust grades for assignments turned in late.	1	2	3	4	5	6	7	8	9

10	I penalize a student's grade for not turning in an assignment.	1 2 3 4 5 6 7 8 9
11	I adjust grades by dropping the lowest or highest score.	1 2 3 4 5 6 7 8 9
12	I adjust grades for students receiving special education services.	1 2 3 4 5 6 7 8 9
13	I adjust grades by weighting some scores higher than others.	1 2 3 4 5 6 7 8 9

Appendix D: Final Draft of Survey Instrument

SURVEY OF GRADING PRACTICES AND TEACHER EFFICACY

Informed Consent

My name is Jason Weaver and I am currently the assistant principal of East Middle School in Joplin, Missouri. As a doctoral student at Southwest Baptist University, I am conducting research regarding the connection between a teacher's grading practices and their level of efficacy.

This study seeks to include all PK-12 teachers in southwest Missouri who are responsible for assigning grades to students. This survey will ask questions regarding your grading practices and level of efficacy as well as a few demographic questions to assist in compiling and analyzing data. The survey should take no more than 15 minutes.

Your participation is completely voluntary. You may chose to withdraw at any time. There is no penalty for not participating or choosing not to answer all questions. All responses are anonymous. No information identifying you individually or your district will be collected, only demographic information used to aggregate results. Responses will be compiled and reported in aggregate only. Your completion and submission of the survey will indicate your consent to participate and for your responses to be included in the study.

This project has been reviewed and approved by the Research Review Board (417-328-1735) of Southwest Baptist University. The committee believes that the research procedures adequately safeguard the subject's privacy, welfare, civil liberties, and rights. For questions about your participation or to receive a copy of the results of the study, please contact me at jasonweaver@joplinschools.org. Thank you for your consideration and time.

Demographic Information

My gender is:

- ☐ Male
- ☐ Female

Which grade-level range best describes the grade(s) you primarily teach?

- ☐ Primary (K-2)
- ☐ Intermediate (3-5)
- ☐ Middle School/Junior High (6-8)
- ☐ High School (9-12)
- ☐ Technical School (9-12)

Which best describes the size of your school district?

- ☐ 1-500 students
- ☐ 500-1,000 students
- ☐ 1,000-5,000 students

- ☐ 5,000-10,000 students
- ☐ 10,000+ students

How many years have you been in education?

- ☐ 1-5 years
- ☐ 6-10 years
- ☐ 11-15 years
- ☐ 16-20 years
- ☐ 20+ years

What is the highest degree you have earned?

- ☐ Bachelor's
- ☐ Master's
- ☐ Specialist
- ☐ Doctorate

Are you directly responsible for assigning grades to students?

- ☐ Yes
- ☐ No

Grading Practices Inventory

The Grading Practices Inventory is designed to gather data about the specific grading practices of individual teachers. To respond to items, consider your own classroom grading and the final grades you issue to students.

What type of final grades do you provide for each student?

- ☐ a single grade for the entire course or subject area
- ☐ separate individual grades for specific competencies, skills or standards in the course or subject
- ☐ both an individual grades for specific competencies, skills or standards as well as a single overall summary grade for the course or subject area

How were your grading policies and practices determined?

- ☐ by personal choice
- ☐ by direction of a building or district grading policy
- ☐ by a combination of building or district grading policy and personal choice

	Grading Practice	Never		Seldom		Sometimes		Quite a bit		A Great Deal
1	I include evidence of academic performance in student grades (journals, tests, homework, teacher observation).	1	2	3	4	5	6	7	8	9
2	I include evidence of non-academic performance in student grades (attendance, behavior, participation, effort, attitude).	1	2	3	4	5	6	7	8	9

3	I assign grades based on student performance against a standard or objective.	1 2 3 4 5 6 7 8 9
4	I assign grades based on student performance against their own past performance (growth or improvement grading).	1 2 3 4 5 6 7 8 9
5	I assign grades based on student performance against overall class performance (curving grades).	1 2 3 4 5 6 7 8 9
6	I determine a final grade by averaging all scores for a given subject, standard or course.	1 2 3 4 5 6 7 8 9
7	I determine a final grade by considering the median, mode, or other statistical measure for all scores for a given subject, standard or course.	1 2 3 4 5 6 7 8 9
8	I determine a final grade by using my own professional judgment.	1 2 3 4 5 6 7 8 9
9	I adjust grades for assignments turned in late.	1 2 3 4 5 6 7 8 9
10	I penalize a student's grade for not turning in an assignment.	1 2 3 4 5 6 7 8 9
11	I adjust grades by dropping the lowest or highest score.	1 2 3 4 5 6 7 8 9
12	I adjust grades for students receiving special education services.	1 2 3 4 5 6 7 8 9
13	I adjust grades by weighting some scores higher than others.	1 2 3 4 5 6 7 8 9

Teacher Sense of Efficacy Scale (Short Form)

Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. Please indicate your opinion about each of the statements below. Your answers are confidential.

Teacher Beliefs		Nothing	Very Little	Some Influence	Quite a Bit	A Great Deal
1	How much can you do to control disruptive behavior in the classroom?	1	2	3 4 5 6 7	8	9
2	How much can you do to motivate students who show low interest in school work?	1	2	3 4 5 6 7	8	9
3	How much can you do to get students to believe they can do well in school work?	1	2	3 4 5 6 7	8	9
4	How much can you do to help your students value learning?	1	2	3 4 5 6 7	8	9
5	To what extent can you craft good questions for your students?	1	2	3 4 5 6 7	8	9
6	How much can you do to get children to follow classroom rules?	1	2	3 4 5 6 7	8	9
7	How much can you do to calm a student who is disruptive or noisy?	1	2	3 4 5 6 7	8	9
8	How well can you establish a classroom management	1	2	3 4 5 6 7	8	9

	system with each group of students?	
9	How much can you use a variety of assessment strategies?	1 2 3 4 5 6 7 8 9
10	To what extent can you provide an alternative explanation or example when students are confused?	1 2 3 4 5 6 7 8 9
11	How much can you assist families in helping their children do well in school?	1 2 3 4 5 6 7 8 9
12	How well can you implement alternative strategies in your classroom?	1 2 3 4 5 6 7 8 9