Electronic Theses and Dissertations

12-1-2016

Teachers' Perception of their Grading Practices by Experience, Training, and State-Assessed Effectiveness

Jeremy David Hinson

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation

TEACHERS' PERCEPTIONS OF THEIR GRADING PRACTICES BY EXPERIENCE,
TRAINING, AND STATE-ASSESSED EFFECTIVENESS

by

Jeremy D. Hinson

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Education

Major: Leadership and Policy Studies

The University of Memphis

December 2016

Acknowledgments

I want to begin by honoring the memory of Dr. Larry McNeal. He made a huge impact on my professional life and is responsible for leading me toward the doctoral route. I would like to thank Dr. Reginald Green for taking me on and becoming my dissertation chair after the loss of Dr. McNeal. Dr. Green's wisdom and guidance was invaluable throughout this process. I would like to thank Dr. Louis Franceschini for guiding me through the development of my survey and chapter 4. I would also like to thank Dr. Charisse Gulosino for taking so much time out of her busy schedule to lead study groups for the stats comp. I want to thank Dr. Gulosino, Dr. Franceschini, and Dr. Boudreaux for agreeing to be on my dissertation committee. I would especially like to thank Dr. Tim Fite for his help through the survey process of this study.

I would like to thank my wife April for her patience throughout this process. She has happily taken on more responsibilities so that I could obtain this goal. I want to thank my parents Jerry and Lois; who instilled in me the value of education at a very early age. I want to thank my daughters Avery and Ally for inspiring me to complete this degree everyday.

I also want to thank my cohort. This was a marathon, and we relied on one another along the way. Kemmashela Brooks Conner, Dr. Tonya Cooper, Dr. Margaret Gilmore, Dr. Angela Hargrave, Tim Jones, Rebecca Scott, Dr., David Stephens. I also want to thank Jody Hinson, Brandi Hinson, Joey Harrison, Eric Ellerbrook, Connie Thompson, Cal Johnson, Jenny Hatch, Seela Newbill, Charlotte Haslett, Misty Odom, Amy Holden, and Valerie Hogue for their input and support throughout this process.

Abstract

Hinson, Jeremy David. Ed.D. The University of Memphis. December 2016. Teachers' perception of their grading practices by experience, training, and state-assessed effectiveness. Major Professor: Reginald Green, Ph.D.

The literature reveals a century-long acknowledgment of inconsistency within the grading process. Although scholars and practitioners acknowledge this inconsistency, teachers' subjective judgments continue to be a large factor when determining students' grades. The literature also highlights practitioners' lack of voice regarding the subject of grades. Through the use of surveys given to K-12 public educators, this research attempts to give voice to these practitioners.

This research examined teachers' perceptions of their classroom grading practices. The researcher was interested in analyzing the relationships between effective teachers' thoughts on classroom assessment. Comparisons were made on teachers' on factors such as teacher effectiveness, training, and years of experience. Approximately 240 teachers were surveyed concerning their perceptions of prevailing grading practices.

Data identifying teachers' perceptions of the prevailing grading practice have been securely collected, studied, and analyzed. The data has been analyzed using Pearson's Chi-Square Goodness-of-Fit test. The Pearson's Chi-Square Goodness-of-Fit test compares an anticipated frequency to an actual frequency. In order to calculate chi-square goodness-of-fit, the researcher must develop the null hypothesis in which no statistical significance exists. The researcher must also determine a significance level, in this case 0.5, to determine if there are any data that render statistically significant results for an alternate hypothesis.

The results of interest have been divided into three sections: highly effective teachers, formal assessments, and effort in grading.

Highly effective teachers believe that students' grades are a reflection of their own instructional effectiveness. These results are similar to the Irreplaceables study conducted in 2012.

Formal assessment training made no discernible difference in teachers' perception of grading. These results are interesting and highlight that higher education curriculum might not have the impact to affect change needed in the grading process.

The majority of participants (79.8%) surveyed tended to agree with the fact that they considered student effort when grading. Researchers have argued that this is a questionable practice. They cite the fact that the perception of student effort is difficult to measure and varies from teacher to teacher.

Table of Contents

List of Tables

Chapter 1

Introduction

The 100-point (A-F) grading system, which was developed by universities well over a century ago, has been a staple in American education. Having dwelled within the confines of this academic assessment structure for the majority of our adolescence, most Americans understand the concepts of this 100-point (A-F) grading system better than the concept of the infield fly rule or the ingredients of apple pie. Lynn Olson (1995), an educational reporter, wrote that the use of grades,

> s one of the most sacred traditions in American education…. The truth is that grades have acquired an almost cult-like importance in American schools. They are the primary shorthand tool for communicating to parents how children are faring. (p. 24)

There are some educational innovators that believe our current grading systems are mathematically imbalanced, unreliable, and invalid measures of student learning (O'Connor, 2012). Marzano (2000) stated, "The answer is quite simple: grades are so imprecise that they are almost meaningless" (p.1). Dressel (1983) said that grades are, "an inadequate report of an inaccurate judgment by a biased and variable judge to the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material." (p.12)

The subjective nature of grading makes it impractical for gauging students' knowledge (Dressel, 1983). What is an A in one teacher's class might be a C in another teacher's class. Many items are assessed in grades other than academic knowledge. This includes things such as: "major exams, compositions, quizzes, projects, and reports,

along with evidence from homework, punctuality in turning in assignments, class participation, work habits, and effort" (Guskey, 2011, p.18). Link (2014) stated, "the large majority of the nation's grading policies perpetuates and even encourages the use of ineffective grading practices" (p.1). An example of this can be found within the grading policy of a Tennessee school district: "a reasonable number of academic points may be deducted from a student's academic grade for failure to submit homework or other assigned academic work on the date specified by the teacher" (Legacy SCS Policy 5014 & 5015, 2012, p.1).

Teachers' assessment procedures vary. The autonomy that teachers utilize when grading creates serious inconsistency from district to district, school to school, and even teacher to teacher. Variance can occur in the design of assessments, the regularity of assessments given, the feedback that the teacher provides, and the grade assigned to the student. Grades are supposed to reflect the level to which students have mastered skills. However, when it comes to assigning grades, inconsistency seems to be the norm (McMillan, Myran, & Workman, 2002).

Background to the Study

This unfavorable view of prevailing grading practices is not a new phenomenon. Researchers have been debating the procedures of the grading system for over a century (Guskey & Bailey, 2001). Durm (1993) reported that relatively early after the inception, the grading system was being debated by Finkelstein (1913):

> When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate scholastic attainment of the pupils or students in these institutions, and when we remember how very great

stress is laid by teachers and pupils alike upon these marks as real measures or indicators of attainment, we can but be astonished at the blind faith that has been felt in the reliability of the marking system. School administrators have been using with confidence an absolutely uncalibrated instrument.... What faults appear in the marking systems that we are now using, and how can these be avoided or minimized? (p.1)

Rinsland (1937) wrote, "when all is said and all studies examined, one is forced to admit the whole grading system is highly subjective, unreliable and unfair" (p. 26).

Statement of the Problem

There is an elephant in the classroom. Researchers have been pointing out the shortcomings of the prevailing grading system for decades. Yet practitioners have fallen silent, and remained static, making little to no changes in the system for over a century. Do practitioners disagree with researchers about the fact that grades are biased and subjective? Why has action not been taken on the part of the educational practitioners? Duncan and Noonan (2007) argued that there is insufficient research on the perspective of practitioners concerning grading. Researchers are clear about the inaccuracies of the current grading system; but practitioners have not been given a voice with which to articulate their insights of the prevailing grading/assessment system. The proposed research will explore practitioners' perceptions of assessment and grading practices.

Research Questions

1. Are teachers equally likely to respond positively to 21 statements concerning grading and assessment practices?

2. Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?

3. Is there a significant difference in the perception of assessment and grading practices between teachers who have had formal assessment training and those that have not had training?

4. Is there a significant difference in the perception of assessment and grading practices between teachers who are more experienced and teachers who are less experienced?

Purpose of the Study

The purpose of this study is to gain insight into practitioners' impressions of the current grading process. The study will also examine what facets of the grading process teachers perceive as important. This will include how these perceptions influence the effectiveness of the teachers.

Definition of Terms

Assessment/Grading Practices. Methods, techniques, or tools that educators use to assess, measure, and record the academic readiness, growth, or needs of students (Abbott, 2014).

Average Growth Index. "This is a measure of student growth across the tested grade levels in a school. This index is a value based on the growth across grade levels (TCAP4-8) or within a grade and subject (3rd Grade TCAP, SAT-10, EOC, ACT, PLAN, and EXPLORE) divided by its standard error. The Average Growth Index allows for a direct comparison of school effectiveness (TVAAS, 2014)."

Experience. For the purposes of this study, teachers will be divided into two categories regarding teaching experience: teachers who had taught more than 15 years and those that had taught fewer than 15 years.

Formal Assessment Training. Collegiate coursework with curriculum focused on student assessment.

Four-Point Scale. Four-Point Scale will be understood as a grading scale in which the range of grades is 1-4, using only whole numbers, with no 9 option for a null grade (a zero) and no options for fractional grades (such as a 2.5 or a 3.75). Fractional grades are still allowable in aggregate when computing overall grade point average under this scale, however. The whole number restriction only applies to individual grades given/received on individual assignments and/or assessments (Reeves, 2004a).

Practitioner. Merriam-Webster defines ("Practitioner", 2015) as, "a person who regularly does an activity that requires skill or practice". For the purposes of this research, educational practitioners are K-12 public school educators that practice the skill of teaching.

Process. Student effort perceived through participation and work habits.

Product. Academic achievement measured through exams, papers, and projects.

Progress. Academic growth measured between two points in time
 (Brookhart,1993).

Standard Grading Scale (Model A). Throughout this study, the term (standard grading scale) will refer to the widely used 0-100 grading scale used in classrooms throughout the country. In the Standard Grading Scale, an A falls within the 90-100 range, a B falls within the 80-89 range, a C falls within the 70-79 range, a D falls within the 60-69 range, and an F falls anywhere between the 0-59 range (Reeves, 2011).

Standard Grading Scale (Model B). In some current schools systems, a variation on Standard Model A is used in which the upper and lower limits of a particular grade range are condensed and the F range is increased. For instance, in one version of Standard Model B, an A is often defined as 94-100, a B as 86-93, and so on. Minor variations on this theme are inconsequential for the purposes of this study; thus, throughout this study, Standard Model B will refer to that aggregate of variations on Standard Model A in which the 0-100 scale is maintained, but the grade ranges limiting.

Student Academic Success. Throughout this study, student academic success will be measured against a number of different variables, but, in general, student academic success will be determined as a function of grades earned/received in direct correlation to actual learning as measured against standardized learning outcomes (Reeves, 2006).

Theoretical Framework

At its inception, the concept of systems thinking was used in engineering and called *hard systems thinking* (Banathy, 1996). The prominent focus of this concept was to study the processes of contemplating and solving problems (Link, 2014). This theory evolved to be known as *organismic systems thinking*. The biologist Ludwig von Bertalanffy was the systems theorist who developed *organismic systems thinking*. He used the concept to challenge Descartes' "scientific method" (Link, 2014).

Descartes' scientific method purposed that individual segments within the system could be inserted in a linear manner and evaluated independently. Bertalanffy (1956) argued that Descartes' method was incorrect. Bertalanffy contended that a system is open and interacts with its proximate surroundings. Bertalanffy's open system theory concluded that the environment effects change in the system, and the system effects

6

change in the environment. This concept has shaped the study of all living systems thereafter (Link, 2014).

The emergence of social systems occurred in the second half of the 20th century. Social systems design is a purposeful creation and manifestation of the future that people collectively desire. It is the realization of what should be. Collectively, people create these social systems for desired outcomes or environments (Banathy, 1996).

Katz and Kahn (1966) wrote:

Social structures are essentially contrived systems. They are made of men and are imperfect systems. They can come apart at the seams overnight, but they can also outlast by centuries the biological organisms, which originally created them. The cement, which holds them together, is essentially psychological rather than biological. Social systems are anchored in the attitudes, perceptions, beliefs, motivations, habits, and expectations of human beings. (p. 33)

Social systems are compiled of a variety of individuals with varying backgrounds that function together for a common goal or purpose (Green, 2010). Some examples of large-scale social systems are laws, currency, government, religion, and schools. Schools are complex social systems. The common goal that everyone is working toward within schools is producing educated people that can contribute to society (Goodlad, 2002).

Green (2010) stated:

[Teachers and principals in healthy systems/schools] realize that schools are social systems comprised of a large number of individuals employed to perform specific functions. Because functions in the schoolhouse are interdependent and interrelated, there has to be an interconnection. The extent to which one individual is able to complete a task in an effective manner is dependent on the cooperation, collaboration,

and often, the extent to which the other individuals complete assigned tasks. Because of the interdependency of the work, individuals have to build relationships sufficient to make strong connections with other individuals in the organization. This connection causes the organization to function effectively. (p. 209)

This common goal creates interconnectedness among individuals and the system. School systems have many stakeholders that have a vested interest in the education of individuals within the society. The education level of future generations relates directly to the welfare of that country's industry, economy, and government.

According to Lunenburg (2010), systems like schools have four kinds of inputs from the environment: physical resources, human resources, financial resources, and information resources. Among Lunenburg's inputs, grades are part of the information resources (Link, 2014). Educators gather, compile, and assess academic inputs or information from students daily. These information inputs are synthesized by educators to produce outputs.

Within the open social system that is education in America, outputs are the realization of objectives attained by school or districts (Lunenburg, 2010). Outputs vary from school to school. They can include success and accomplishment or failure to meet expected progress. Grading systems can be used to convey the academic achievement and gains of individual students and the entire school. Grades are used to report academic proficiency to students, parents, and other academic institutions. In spite of the subjective nature of the grading process, this has been the way education reports academic achievement for generations.

Heifetz, Grashow, and Linsky (2009), when discussing change within a social system, wrote, "the reality is that a social system…is the way it is because the people in

that system…want it that way" (p. 17). In a social system, information can be processed and procedures can be performed at the subconscious level (Heifetz et al., 2009). Sometimes members within an organization can become so accustomed to the status quo, that they do not see the need for change.

Significance of the Study

All of the social norms that we abide by, outside of those dictated by physiological needs, are models created by human beings. These models are ingrained so deeply that we often adhere to these norms collectively without giving them much thought. The grading practice predominantly used throughout the United States is a good example of an ingrained social norm. The results of this research could add to a body of existing research about grading practices.

Limitations and Delimitations

This study is limited by the fact that data are gathered solely from a perception survey given to teachers. Perhaps more accurate and comprehensive gathering of information would have occurred if the researcher could have interviewed teachers and/or observed classroom practices. Rather than being self-reported by the teacher, it would have been ideal if teachers' Tennessee Value-Added Data could have been secured directly from the school district. Teachers could have given the researcher an inaccurate account of his/her effectiveness on the survey.

Chapter Overview

Over the years, researchers have offered suggestions to improve the prevailing grading system. Issues with the grading system are multifaceted. The solutions proposed by researchers cover a wide variety of topics. However, grading practices have not evolved over the last century because practitioners are bound by the parameters of the

reporting system that is currently in place. This study will attempt to gain insight into

practitioners' impressions of the current grading system.

Chapter 2

Review of Literature

Introduction

  While grading students may seem a simple task, it is actually a complex process with many variables to consider. What information is significant or worthwhile to impart to students? How will the teacher assess the learning of his/her students (homework, daily grades, exam grades)? How will the teacher provide feedback of student performance to students, parents, and higher education institutions? All of these variables are influenced by teachers' values and beliefs. The validity of grades is completely reliant on the grading practices that teachers adopt (Link, 2014). These practices vary from state to state, district to district, and classroom to classroom.

  This chapter will chronicle the prevailing grading process within K-12 public education in America. The literature will highlight the history of the system. Imperfections within the grading system are vast and multifaceted. The solutions purposed by researchers cover a wide variety of topics. Literature, in this chapter, will document several researchers' observations as well as suggestions for change.

History of the Grading System

  The origin of the current grading system utilized in K-12 education can be traced to the university setting more than a century ago. In all probability, the first collegiate grades given in the United States were at Yale University in 1783 (Durm, 1993). Ezra Stiles, the president of Yale in the late 18th century, documented a 4-point grading scale used to assess 58 students taking an examination: (a) Optimi, (b) Second Optimi, (c) Inferiores (Boni), and (d) Pejores (Durm, 1993). In 1877, more than 200 years after its

1636 inception, Harvard began using the 100-point scale below to assess students (Durm, 1993):

Division 1: 90 or more

Division 2: 89 to 75

Division 3: 74 to 60

Division 4: 59 to 50

Division 5: 49 to 40

Division 6: below 40

In 1895, Harvard adopted three classifications for merit: Failed, Passed, and Passed with Distinction. That same year, Michigan adopted the following: Passed, Incomplete, Conditioned, Not Passed, and Absent. Finally, in 1897, Mount Holyoke University developed the grading system that proved to be the foundation of K-12 educational assessment for the last century:

A Excellent 95-100

B Good 85-94

C Fair 76-84

D Passed 75

E Failed Anything below 75

Prior to adopting this assessment system, teachers in K-12 education largely conveyed student progress to parents directly. This was a narrative description of progress that largely took place in the form of home visits (Nava & Loyd, 1992). In the 19th century, students of all ages were placed together in one-room schoolhouses. Many

students did not attend beyond an elementary school level (Nava & Loyd, 1992). Schools were locally funded, and attending school was not a legal requirement.

Early in the 20th century, lawmakers passed compulsory attendance laws at the elementary level (Guskey, 1994). During this time, high school attendance began to grow exponentially. The number of public high schools in the United States increased from 500 to 10,000 between 1870 and 1910 (Gutek, 1986). High school populations became diversified, while high school curriculum became more subject specific and streamlined (Guskey & Bailey, 2001). During this era, high school teachers began to report students' progress in these specific subject areas using markings similar to those of university professors. These humble markings evolved into the universal grading system that we have today (Kirschenbaum, Simon, & Napier, 1971). The transition to our current grading system was gradual and met little resistance from American educators (Guskey & Bailey, 2001).

One explanation for the ease of this transition to percentage grading is the increased numbers of students in the classroom by the 1910s. Educators seemed to prefer the ease of demonstrating student progress through grades rather than verbally conveying progress to parents. However, in 1912, Starch and Elliot published a study that challenged the subjective nature of the newly formed grading system. These researchers devised an experiment where teachers from 142 different schools graded an identical English paper. Starch and Elliot (1912) found that subjective factors affected teacher grading. Also, punctuation, spelling, and neatness caused variances in grading. In an effort to move away from subjective scoring practices, many educators of this era

were proponents of the idea that there was a natural distribution of intelligence (Guskey & Bailey, 2001).

Educators were using grades to classify and separate students based on academic ability levels. Zirkle (1920) stated, "it is evident in every school the incapable are a drag on the class and the exceptionally bright ones are held back" (p. 189). The notion of a natural distribution of intelligence led many educators of the time to assign their students grades based on a bell-shaped curve. Russell (1930) patented a grading scale that was said to be the best mathematical match for the normal bell-shaped curve. The patent was: 3% A; 23% B; 48%C; 23% D; and 3% E. This process seemed reasonable at the time because people assumed that a student's natural intelligence would correlate to their academic achievement (Middleton, 1933).

The 1920s brought about the progressive education movement. Proponents of this movement encouraged the discontinuation of the uncompromising grading system. They used new findings from social sciences to justify the focus of the individual child (Cohen, 1974). Citing the fact that formal grading hindered the teaching and learning process, some schools chose to remove the grading system altogether (Chapman & Ashbaugh, 1925). Some schools elected to revert back to the verbal assessment of students' academic progress (Good, 1937). Other schools insisted that students master specific curriculum before moving on to other concepts (Heck, 1938).

E. George Payne, an educational sociologist of the 1930s, advised educators of this era to modify assessment practices in order to prevent inequalities to students (Rinsland, 1937). Warren Middleton was selected to lead a committee to review the

14

grading process within this school. The following is how he described embarking on this experience:

> The Committee On Grading was called upon to study grading procedures.
> At first, the task of investigating the literature seemed to be a rather hopeless one.
> What a mass and what a mess it all was! Could order be brought out of such
> chaos? Could points of agreement among American educators concerning the
> perplexing grading problem actually be discovered? It was with considerable
> misgiving and trepidation that the work was finally begun. (Middleton, 1933, p. 5)

By the 1940s, educational participation among high-school-aged Americans had moved toward becoming the norm. Attendance among high-school-aged students increased from 6%-7% in 1890 (Kliebard, 2004), and to 73% by the 1940s (U.S. Census Bureau, 1993). Early in this decade, the world's attention was transfixed on the European and Pacific theaters. Soon after the completion of World War II, critics again turned their focus back to the negative issues with grading. Wrinkle (1947) urged educators to use grades as a formative assessment instrument within a supportive environment, rather than a summative reflection of one's effort.

In 1957, the Soviet Union launched the Sputnik satellite. Americans looked to the stars to witness a man-made contraption flying overhead. Some took issue with the fact that this flying machine was not built by Americans. Webster (2011) wrote, "new arguments framed assessment and grading in the context of national security" (p.20).   In the midst of the Cold War, Americans found themselves in somewhat of an intellectual competition with the Soviet Union. Soon President Kennedy initiated the space race. In a speech, John F. Kennedy (May 25, 1961) stated, "I believe that this nation should

commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to earth". As cited in Webster (2011), this competition renewed the use of grades to track and ability group American students

In 1971, the National Council of Teachers of English approved a policy on grading that progress monitored academic growth by using teachers' observations, comparative work samples, and anecdotal records. The council's policy placed less emphasis on letter grades as an indicator of student progress. Webster (2011) wrote, "The policy reflected the growing concern about failure and its stigma" (p. 21).

In 1983, the National Commission on Excellence in Education released *A Nation at Risk*. This study compared the students of the United States with other students around the world. Some questioned the validity of this study. The reaction to this study prompted many to advocate for national education standards (Webster, 2011).

In 2001, President George W. Bush's No Child Left Behind (NCLB) initiative held schools accountable in a way not previously imagined (Fritzberg, 2001). NCLB mandated that all states develop achievement examinations, which were aligned with NCLB standards, to measure student's academic abilities in Grades 3-8 annually. While the federal government left it up to each individual state to determine the proficiency level of its annual examination, each state was required to establish individual performance standards for all schools in the state with a 100% passing rate in reading and math by 2014 (Fritzberg, 2004). Each state set achievement goals and communicated the consequences for not meeting those goals to all the districts within the state. If schools did not meet Adequate Yearly Progress (AYP) they were labeled:

Year 1 "Target"

Year 2 "School Improvement 1"

Year 3 "School Improvement 2"

Year 4 "Corrective Action"

Year 5 "Restructuring"

The intensity of sanctions increased with each year a school did not meet AYP. Eventually, federal funding was withdrawn from the school and personnel were replaced (Fritzberg, 2004).

In 2010, authorized under the American Reinvestment and Recovery Act of 2009, the Race to the Top Assessment program provided substantial funding from the federal government to states that would make an effort to improve instruction, measure student achievement against rigorous curriculum, and develop consistent assessments. In order to qualify, these assessments had to

produce data (including student achievement data and student growth data) that can be used to inform (a) determinations of school effectiveness; (b) determinations of individual principal and teacher effectiveness for purposes of evaluation; (c) determinations of principal and teacher professional development and support needs; and (d) teaching, learning, and program improvement. (United States Department of Education, 2010 p. 4)

Also, for states to qualify for Race to the Top money, they needed to adopt college and career ready standards. This initiative would later be called Common Core standards (Strauss, 2014). The Common Core State Standards are a transparent set of shared objectives and expectations for the knowledge and skills students in kindergarten

17

through 12 grades need. These academic standards are projected to outline the learning

needed in English language arts and mathematics. Mastery of these shared standards

prepares students to enter the workforce and/or introductory college courses.

This initiative was developed in 2009 by the National Governors Association

Center for Best Practices and the Council of Chief State School Officers. Leaders from

48 states originally took part in the development of the Common Core Standards. At one

time, 45 states were taking part in some aspect of the Common Core process (Common

Core State Standards Initiative, 2015).

Politics have taken a toll on the Common Core State Standards. Those that are

opposed to Common Core Standards argue that the federal government is obstructing

state's rights. Controversy over the uniformity of this process has ensued (Baker, 2013).

Oklahoma and South Carolina were the first two states to repeal Common Core Standards

that they had previously adopted (Bidwell, 2014; see Figure 1). Other states followed

suit.
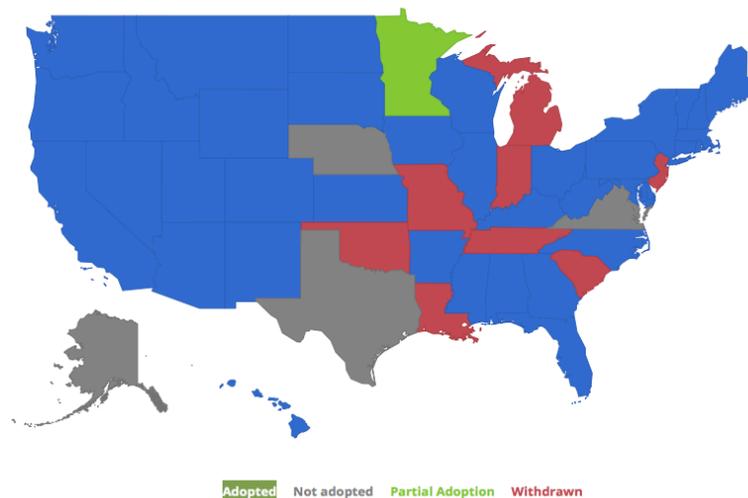
Figure 1. Common Core Website

The Purpose of Grading

Grades can be used in two distinct ways. The first is to report academic proficiency to students, parents, and other academic institutions. In this way, grades are used as summative assessments of student performance over time. Parents, higher education admissions boards, armed forces recruiters, or potential workplace personnel directors can formulate opinions about a student's academic competence based on these summative assessments.

The other way grades can be used is as formative assessments. Tomlinson (2005) concluded that grades should fundamentally be used to provide feedback to students. This feedback allows students to correct deficiencies and explore subject matter more thoroughly. Guskey (1997) encouraged teachers to provide specific suggestions when grading student work. Researchers have documented the value of specific feedback in the learning process (Black & William, 1998; Hattie & Timperley, 2007).

Points of Agreement

Practitioners are bound by the parameters of the reporting system that is currently in place. Despite the many discrepancies within this system, Guskey (1996) highlighted several points of agreement that exist in the grading system:

1.   Grading and reporting are not essential to instruction. Teachers do not need grades
        or reporting forms to teach well, and students can and do learn without them. We
        must recognize, therefore, that the primary purpose of grading and reporting is other
        than facilitation of teaching and learning.

2.   No one method of grading and reporting serves all purposes well. Various grading
        and reporting methods are used to: (a) communicate the achievement status of

students to their parents and other interested parties; (b) provide information to students for self-evaluation; (c) select, identify, and group students for certain educational paths or programs; (d) provide incentives for students to learn; and (e) document students' performance to evaluate the effectiveness of instructional programs.

3. Grading and reporting will always involve some degree of subjectivity. Regardless of method used, assigning grades or reporting on student learning is inherently subjective. In addition, the more detailed the reporting method and the more analytic the process, the more likely subjectivity will influence results.

4. Grades have some values as rewards, but no value as punishments. Although educators undoubtedly prefer that motivation to learn be entirely intrinsic, grades and other reporting methods are important factors in determining how much effort students put forth. Most students view high grades as positive recognition of their success, and some work hard to avoid the consequence of low grades. At the same time, no studies support the use of low grades as punishments. Instead of prompting greater effort, low grades most often cause students to withdraw from learning.

5. Grading and reporting should always be done in reference to learning criteria, and never on the curve. Using the normal probability curve as a basis for assigning grades typically yields greater consistency in grade distributions from one teacher to the next. The practice, however, is detrimental to both teaching and learning. Grading on the curve communicates nothing about what students know and are able to do, and grading on the curve makes learning a highly competitive activity in which students compete against one another for the few scarce rewards (high grades) distributed by the teacher. When grading and reporting relate to learning criteria, teachers are able to provide a clearer picture of what students have learned. There are three broad

learning criteria categories: product, process and progress criteria. (p. 28)

Grading Criteria

What criteria are being used to assess students in the prevailing grading system? Brookhart (1993) studied three types of grading criteria used by teachers:

1. Product (Academic achievement measured through exams, papers, and projects etc.)

2. Progress (Academic growth measured between two points in time)

3. Process (Student effort perceived through participation and work habits)

Product (achievement) criteria. A grade is supposed to provide a precise, undiluted indicator of a student's mastery of learning standards (Wormeli, 2006). Multiple studies suggest that teachers believe achievement is the most important aspect of the grading process (Allen, 2005; McMillan, 2007; O'Connor, 2009; Tomlinson, 2005). Guskey (2006a) wrote, "Teachers and students alike prefer this approach because they consider it both fair and equitable" (p. 672). However, Brookhart (2009) found that teachers refrained from using achievement for grades exclusively because of their concerns about student motivation and self-esteem.

Product (achievement) limitations. The problem with using academic achievement exclusively as a yardstick for measuring teacher, administrator, school, or district effectiveness is that academic achievement correlates with socioeconomic status of children (Kennedy, Peters, Thomas, 2012; see Figure 2).

Economic Disadvantage vs. Student Achievement

Economic Disadvantage vs Student Achievement
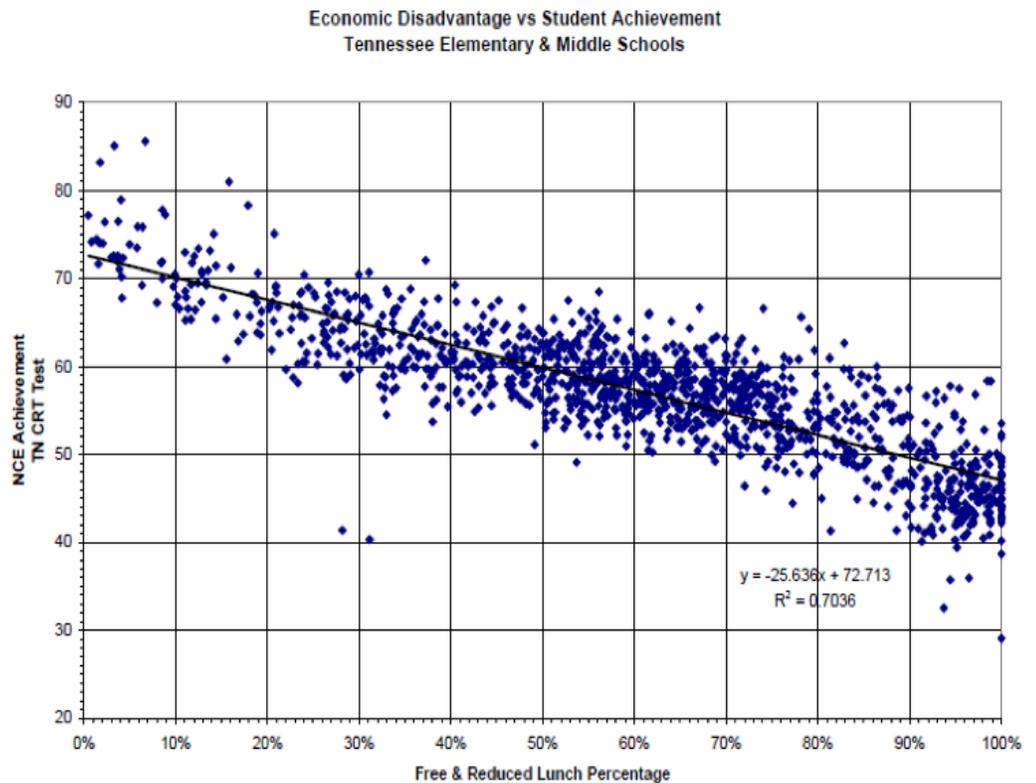Tennessee Elementary & Middle Schools

Figure 2. Battelle for Kids, 2010

Educational shifts toward academic accountability, like No Child Left Behind, have shone a glaring light on the achievement gap. An example of this can be taken using data from the yearly achievement tests given during Tennessee's 2012-13 school year. The Tennessee State Report Card shows that 49.2% of students scored below

proficient levels on the math examination in Grades 3 through 8. This equates to approximately 488,000 students throughout the state (Tennessee Department of Education, 2013). If all things were equal, grading on students' mastery of learning objectives would be ideal. The societal issues causing the achievement gap are at the heart of the grading debate.

Progress (growth) criteria. Agnew (1985) conducted a survey that found teachers perceived progress criteria to be the second most important aspect of grading. In assessing progress, teachers measure how much information a student learns between two points in time as the primary determinate for grade consideration. At the classroom level, some researchers have suggested that teachers use progress criteria as a portion of students' grade (McLoughlin & Lewis, 2008; Silva, Munk, & Bursuck, 2005).

At the state level, Value-Added Accountability Systems (VAAS), which were developed by Dr. William Sanders, analyze data from individual students over time. The data used for VAAS come from yearly achievement tests. The system has the capability to report individual student scores compared to others in the district, state, or nation (achievement scores), and the system also has the ability to allow each student to act as his or her own control comparing how much academic growth they had made in any given year (student growth data; Goldstein, 2005). The value-added measurements are being used to measure not only student growth, but also school and individual teacher effectiveness.

Progress (growth) limitations. When considering the progress of a student at the classroom level, a teacher must consider the ability of that student. This consideration is often subjective in nature. Also, because the growth of each student is so

individualized, reporting the grade can become difficult and confusing. What constitutes an A for one student, based on his/her ability level and potential for growth, is completely different for another student (Guskey, 2006a).

Process (effort) criteria. Process criteria can include student participation, effort, and work habits (Guskey & Jung, 2009). Teacher perception of these factors can influence students' grades (Zoeckler, 2007). Guskey (2006a) wrote, "Many teachers point out that if they use only product criteria in determining grades, some high-ability students will receive high grades with little effort, while the hard work of less-talented students will go unacknowledged" (p. 672). Teachers often factor effort in the case of underachieving students (Stiggens, Frisbie, & Griswold, 1989) because "these students find the relationship between high effort and low grades frustrating and often express their frustration with indifference, deception, or disruption" (Guskey, 2006, p. 672).

Process (effort) limitations. Some researchers recommend that the use of nonacademic measures be excluded from the grading process altogether. Stiggens and Knight (1997) called the factors within the process criteria "grade pollution" (p.61). Because of the subjective nature of effort, it is difficult for parents, teachers, and schools to fully grasp the extent to which student learning has taken place.

Reporting separately. Traditionally, teachers have taken portions of assessment aspects from each criteria (product, progress, and process), and reported them under one letter grade. Guskey (2006) and other researchers encouraged practitioners to continue using the three criteria, but to report them separately (Stiggins, Arter, Chappuis, & Chappuis, 2004).

Reporting separately (limitations). The technique of reporting grades

separately would be extremely difficult for an individual teacher to effectively implement. The decision to report grades separately would need to be a school or district decision and involve educating the teachers, parents, and students regarding the practice. Also, reporting grades separately does not change the subjectivity found within the grading process. It essentially provides parents a transparent account of the individual variables that teachers use to assess student learning.

      Final grading decisions. Despite its limitations, most teachers today use all three of Brookhart's (1993) criteria to determine final grades. Each teacher utilizes autonomy when collecting evidence of student growth and achievement (Cizek, Fitzgerald, & Rachor, 1995. This evidence can be taken from presentations, projects, quizzes, writing tasks, tests, and/or demonstrations (Guskey, 2002b). This teacher autonomy leaves a large variation on the final grade depending on what factors are seen as most important to the teacher.

Teacher Perception/Beliefs

      The factors that are most important to teachers play a major role in developing many classroom procedures. Rubie-Davies, Flint, and McDonald (2011) noted that teachers' beliefs guide their actions, and teachers' actions affect student academic growth and achievement. Though extremely important in the process of learning and teaching, Rubie-Davies et al (2011) pointed out that research involving teacher beliefs are far less examined compared to students' beliefs. McMillian (2007) argued that research on teacher beliefs should expand. Comprehending how teacher beliefs impact teaching practices, favorably or negatively, could result in a breakthrough in student learning outcomes (Koloi-Keaikitse, 2012).

      Teacher beliefs are often viewed with differing perspectives because of its

complexity (Koloi-Keaikitse, 2012). Oliver and Koballa (1992) conducted research asking teachers to define "teacher beliefs." In the study, some teachers offered psychological responses such as, knowledge, values, and attitudes. Other teachers viewed "belief" an actionable concept that influenced behaviors, attitudes, and practices. Other researchers believe that teacher beliefs are too complex to define and cannot be quantified (Cantu, 2001). Koloi-Keaiktse (2012) stated that the complexity of teacher beliefs has led researchers to classify teachers into three categories based on their beliefs about classroom practices: (a) realist teachers; (b) contextual teachers; and (c) relativists teachers.

Realist teachers prefer assessments in which students choose an answer bank rather than developing answers on their own (Nitko, 2001; Segers & Dochy, 2001). These assessments are focused on memorization. Students are expected to learn and retain information over the short term (Segers & Dochy, 2001). Realist teachers also prefer norm-referenced tests. In norm-referenced testing, students' mastery of core knowledge and skills is evaluated relative to the performance of others (Koloi-Keaiktse, 2012).

Contextual teachers often use alternative assessments that are performance based. These teachers use group work, portfolios, and projects to deliver curriculum and obtain grades. Contextual teachers devise assessments that force students to demonstrate a deeper knowledge of content being taught. These assessments could include extended writing and problem solving skills (Haladyna, Downing, & Rodriguez, 2002).

Relativists make up the final subgroup of teachers. The foundation of this assessment practice is found within developmental theory. These educators believe that students thrive in settings that are developmentally appropriate. Relativist teachers assess students in a variety of ways such as oral presentations, technological presentations, written tests, and media (Koloi-Keaiktse, 2012).

Some researchers believe that these teacher beliefs frame what is important to

teachers, thereby guiding actions and procedures. Bliem and Davinroy (1997) stated that teachers' truths and beliefs regarding classroom procedures are developed both by experiences they have had as students and as teachers. These deeply held beliefs determine how pliable teachers will be when confronted with educational changes that relate to best practices (Bliem & Davinroy, 1997). These beliefs

> can serve to facilitate or hinder teachers' efforts as they set about altering their actions in the classroom, depending on whether and the extent to which their existing beliefs overlap with the philosophical underpinnings of proposed changes to their practices. (Link, 2014, p. 39).

It stands to reason that the subjective nature of grading is a by-product of teacher beliefs. The following list highlights the latitude given to each individual teacher:

- The choice of curriculum being taught;

- Assignments to reinforce learning;

- How those assignments are disseminated to students;

- How to decide if/when each student displays mastery of academic concepts;

- Communicating to the student and his/her parent academic progress; and

- Develop formative and summative assessment that are fair and accurate to all students

The way that teachers implement all of the aforementioned tasks depends largely on their beliefs of classroom procedures and grading practices (Bliem & Davinroy, 1997). "Thus, if researchers desire to increase student classroom success by altering assessment and grading practices, we must understand the belief system, or perceptions, underlying teachers' ways of evaluating their students' learning" (Link, 2014, pp. 39, 40).

Proposals for Change

   Recommendations for the Elimination of Grades. Chapman and Ashbaugh (1925) cited concerns about the validity and reliability of grades. These researchers recommended the elimination of the formal grading system in the 1920s. Kohn (1993) also alluded to the elimination of grading by expressing concern that grades were harmful motivators for students. Brookhart (2009) stated, "in a perfect world there would be no grades – at least not as we know them now" (p. iii).

   Recommendation for the Recalculation of Grades.   Traditionally, teachers have assigned grades from tests, quizzes, and assignments in a grade book. These grades have been recorded and then averaged to calculate a grade. Researchers have suggested that grading practices could be more effective by eliminating some practices currently used in grade calculations. These suggestions include (a) eliminate the practice of averaging points, (b) end the use of zeros, (c) distinguish formative assessment data from summative assessment data, and (d) replace normative grading with criterion referenced grading (Webster, 2011).

   Some researchers have recommended the discontinuation of the practice of averaging points to obtain final grades. Averaging grades penalizes students for work done in the past and lessens the impact of academic growth and improvement (Airasian, 2005). Zoeckler (2007) stated that averaging grades over time could interfere with the truthfulness of learning that has taken place. McTighe and O'Connor (2005) encouraged practitioners to replace old evidence with new evidence to determine students' final grades.

Assigning students an averaged grade with the mathematical effect of "0" is a misrepresentation of student learning (Guskey, 2006). "Imagine three math tests, two with 100% accuracy, and one not taken and given a 0%. The average grade of 66%, a D, is an inaccurate report of a student's mathematic understanding" (Webster, 2011, p. 51). This inaccuracy and misrepresentation of student learning cause some researchers to recommend the discontinuation of giving students zeros (McMillan, 2007; O'Connor, 2009). Guskey (2000) said that

> teachers defend the practice of giving zeros because students cannot be given credit for work that was not completed: Students certainly should learn to accept responsibility for their actions and should be held accountable for their work. Nevertheless, no evidence demonstrates that assigning zeros helps teach students these lessons. An alternative approach is to assign an I (or Incomplete) grade with explicit requirements for completing the work. For example, students whose work is incomplete or not turned in on time might be required to attend after school study sessions or special Saturday classes until their work is completed to a satisfactory level. In other words, they are not let "off the hook" with a zero. Instead, students learn that they have certain responsibilities in school and that their actions have specific consequences. Not completing assigned work on time means that students must attend special after school sessions to complete the work. Implementing such a policy may require additional funding and support; still, the payoffs are likely to be great. Not only is this approach more beneficial to students than simply assigning a zero, it is also a lot more fair. (p. 27)

The majority of measurement specialists caution against the use of grades for evaluating or sorting students. The idea that only a certain percentage of students can be successful is inappropriate (Webster, 2011). Other researchers call for teachers to end the use of "bell curve" practices. Normal bell-shaped curve explains the distribution of randomly occurring phenomenon when nothing intervenes (Guskey & Bailey, 2001). Tyack and Tobin (1994) contended that educational change in "real school" is extremely difficult even when changes will clearly result in positive results.

Recommendations for Grades as Feedback. Fundamentally, the purpose of grading should be to provide feedback to students regarding learning (Tomlinson, 2005). Students could benefit from specific suggestions for improvement regarding their learning (Guskey, 2006b). Teacher feedback to students has been documented to improve learning and academic performance (Black & William, 1998; Hattie & Timperley, 2007). Page's (1958) study, which included 74 teachers, found that there was substantial academic improvement in student learning when teachers included comments with their grades.

Recommendations for Comprehensive Systems. Many researchers believe there are too many variables within the learning process to comprehensively report learning with a single reporting mechanism (Allison & Friedman, 1995). Instead of relying on report cards, some researchers recommend a comprehensive reporting system (Guskey, 2002). Guskey (2002) envisioned a reporting system including the following items: report cards, narratives attached to reports cards, test results, phone communication, progress reports, school open houses focusing on assessment procedures, personal correspondences, portfolios of student work, exhibits of student work, and

30

conferences (Webster, 2011). Over the years, researchers have offered suggestions to
improve the prevailing grading system. Issues with the grading system are multifaceted.
The solutions proposed by researchers cover a wide variety of topics. Proposals for
Grade Change Table highlights some proposed suggestions for improving grading
policies.

Table 1

Proposals for grade change

Proposals for Change	Researchers	Brief Explanations
Elimination of grades	Chapman & Ashbaugh (1925) Kohn (1999)	Argued against the subjective nature of grades.
Elimination of grades	Willis (1993)	Argued that grades are not only unreliable, but also harmful to students.
Averaging grades	Airasian (2005) Zoeckler (2007) McTighe & O'Conner (2005) O'Connor (2009) Marzano (2000)	Discontinue the practice of averaging points to attain a grade. These researchers argued that averaging grades penalizes students for past assignments.
		They suggested that teachers replace old assignments with new in order to determine the final grade.
Giving zeros	Guskey (2006b) McMillan (2007) O'Conner (2007) Wormeli (2006)	These researchers pointed out the mathematical imbalance of the 100-point grading system. They suggested eradicating the practice of giving zeros.

Table One (continued)

Proposals for Change	Researchers	Brief Explanations
Bell curve	O'Conner (2009) Guskey & Bailey (2001) Wiggens (1993)	These researchers call for teachers to end the use of "bell curve" practices. "Normal bell-shaped curve describes the distribution of randomly occurring events when nothing intervenes" (Guskey & Bailey, 2001).
Providing specific feedback	Guskey (1997) Black & William (1998) Hattie &Timperley (2007)	Rather than just a letter or number grade, research has shown that students benefit from specific feedback from teachers on assignments.
Teacher training	Stiggins (2000) McTighe & O'Conner (2005)	Teacher training on proper assessment practices can lead to a decreased bias and improved consistency in grading practices.
Administration collaborating with teacher on assessment practices	Marzano, Waters, and McNulty (2005)	These researchers found that effective administrators collaborated with teachers on "assessment activities at the classroom level".

Table One (continued)

Proposals for Change	Researchers	Brief Explanations
Debate over product assessment criteria	Airasian (2005) Bigham Baron (2000) Brookhart (2009) Cizek (1996) Covington (2004) Frary (1993) Guskey & Bailey (2001) Linn & Gronlund (2000) Marzano (2001) O'Conner (2009) Popham (2004) Reeves (2006)	**Product Criteria:** Evidence based on achievement of academic standards. A summative evaluation is given to students to assess their mastery of the preset academic standards. This assessment technique is the most agreed upon by researchers.
Debate over process assessment criteria	Ebel & Frisbie (1991) Stiggins (1989) Guskey & Bailey (2001) Marzano (2000)	**Process Criteria:** Many researchers suggest that teachers discontinue the practice of grading based on their perception of students' effort, participation, and motivation.
Debate over progress assessment criteria	Ames and Ames (1991) Wiggins (1996)	**Progress Criteria:** These researchers advocate the use of growth data (how much information a student learns between two points in time) as a measurement for grading consideration.

Effort in Grading

Researchers have urged against the use of process criteria (effort) in grading for

years. Marzano (2000) argued, "…it is appropriate to provide feedback to students on their effort, behavior, and attendance; ideally this feedback should be kept separate from that provided on academic achievement" (p. 39). Stiggins and Knight (1997) wrote that in terms of effort, "definitions of trying hard vary greatly from teacher to teacher" (p. 418). Finally, O'Connor (2002) said, "strong effort, active participation, and positive attitude are highly valuable attributes, but they are reporting variables, not grading variables" (p. 100).

Despite the negative aspects of focusing on process criteria (effort) in grading, it has proven difficult for teachers to subside from this practice. The notion that we are to pass along a positive work ethic to future generations is deeply rooted in our culture. In a teaching manual from 1917, Woofter cited five behaviors teachers were to instill in their students: promptness, industry, politeness, order, and self-control (Webster, 2011). Woofter's century-old guidance to teachers remains relevant in the practices of some modern teachers while grading students. Stiggins et al. (1989) reported the majority, while Blount (1997) reported 80% of teachers interviewed used effort to determine grades in their high school classrooms. Many studies (Agnew, 1985; Johnson, 2001; Wiggins, 1994) have used teacher survey answers to indicate that pupil work habits, behavior, and effort were used to determine grades.

Brookhart (1993) reported that teachers find merit in student effort because it promotes student engagement in the classroom. Some educators insist that low grades motivate students to try harder in order to avoid future poor grades. When using grades for this purpose, some educators believe that receiving high grades serves as positive reinforcement and stimulates students to work hard (Webster, 2011). Grades can be used

to manipulate student behavior (Kohn, 1999). Link (2014) wrote, "teachers will bribe students with good grades (A's) or threaten students with poor ones (F's) to complete assignments" (p. 29).

These assignments that teachers "bribe" and "threaten" students with do not often take the form of major assessments. Guskey (2006a) labeled teachers using the process (effort) method as those who are calculating quizzes, homework, and punctuality of assignments into the grading process. These teachers believe grades should reflect not only the final results but also how students got there. (Guskey, 2006a) How students "get there" is the fundamental process of learning.

Learning is acquiring knowledge or skills through examination, instruction, or experience. In most cases, learning requires the student's effort through participation and positive work habits. Of Brookhart's (1993) grading criteria, product (achievement) and progress (growth) are actual academic measures. The process criteria (effort), however, are the subjective assessment of student effort through classroom engagement and practice. Teachers use it to motivate students in completing assignments to obtain knowledge (learn). Would students possess the intrinsic motivation to complete assignments, which enhance their learning, without grades? (Link, 2014)

Conclusion and Summary

The literature reveals a century-long acknowledgment of inconsistency within the grading process. Although scholars and practitioners acknowledge this inconsistency, teachers' subjective judgments continue to be a large factor when determining students' grades. Another component highlighted within the literature are practitioners' lack of voice regarding the subject of grades.

Chapter 3

Methodology

Introduction

The purpose of this study is to gain insight into practitioners' impressions of the current grading process. The study also examined what facets of the grading process teachers perceive as important and the degree to which they agree about what is important. The researcher intends to determine if there is a significant difference in the perception of assessment and grading practices in highly effective teachers, teachers with more experience, and teachers that have received formal assessment training. The following research questions drive this study.

1. Are teachers equally likely to respond positively to 21 statements concerning grading and assessment practices?

2. Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?

3. Is there a significant difference in the perception of assessment and grading practices between teachers who have had formal assessment training and those that have not had training?

4. Is there a significant difference in the perception of assessment and grading practices between teachers who are more experienced and teachers who are less experienced?

This chapter will discuss the design of the research, the instrumentation, how the data was collected, and the demographic information for the participants of the study.

Research Design

The research design symbolizes the paradigm through which the dissertation can be viewed. The research design guides the data collection and interpretation processes (Creswell, 2008). Principally, research design can be divided into 3 categories: quantitative, qualitative, or "mixed" methods (Ayiro, 2012, Creswell, 2008). There are many adaptations among these 3 categories.

This study investigates the relationship among variables. Because it could also be described as numeric rather than narrative, this study is an example of quantitative research. The current study could further be described as "correlational" because the factors it seeks to investigate are specifically selected and sought after. Correlational studies attempt to establish relationships between variable using statistical measures of connection and difference.

Lastly, the current research could be called survey research because it collects data from a series of pre-determined answers to a set of pre-determined questions. Survey research targets "current attitudes, beliefs, opinions, or practices" (Creswell, 2008, p. 389) and produces item frequencies. The researcher chose a survey design to collect, describe, and compare teachers' perceptions of grading. Surveys can be a valuable instrument for gathering data regarding human characteristics such as perceptions, thoughts, and actions (Dillman, Smyth, & Christian, 2009). Surveys are especially useful, in the field of education, because they help delineate information from large sample sizes and streamline the views and beliefs of many professionals (Burstein, 1978).

Demographics

The participants that took part in this survey were from 3 public school districts in West Tennessee. This population was made up of elementary (grades K-5), middle

(grades 6-8), and high school (grades 9-12) teachers from public schools. These were full-time teachers with a minimum of a bachelor's degree. The survey revealed that 87.55% of the survey participants were female. Male participants made up only 12.45% of those surveyed.

Instrumentation and Validity

This study is an extension of a previous study that used secondary data taken from a public school district in southwest Tennessee. This secondary data was obtained through a survey, adapted by Dr. Laura Link from Liu (2006), called the Teachers' Perceptions of Grading Practices (TPGP). This survey/questionnaire obtained results from approximately 3,000 teachers concerning their perceptions of prevailing grading practices. Teacher responses were grouped together by the school they work.

Liu's (2006) survey was developed and validated for the purpose of assessing teachers' perceptions. This survey is called the Teachers' Perceptions of Grading Practices (TPGP). This questionnaire measures teachers' perceptions of prevailing grading practices in six sections. Questions that rendered the most statistically sound, in previous research results, were taken from (TPGP) and employed to develop the survey used in this current research. A copy of the survey used for the current research, which has been vetted and approved by Dr. Louis A. Franceschini, III, Ph.D, can be found in the appendix.

In order to complete the survey used in this study, participants are asked to click on the answer that most accurately reflects their perception with responses ranging from strongly disagree to strongly agree based on a 5-point Likert rating scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree).

The Tennessee Value-Added Accountability System (TVAAS) was used by the researcher as a measure of academic success for teachers. The system, which was developed by Dr. William Sanders, analyses data from individual students over time. The data used for TVAAS come from yearly achievement tests given for students in Grades 3 through 8 in five subject areas: math, reading, language arts, social studies, and science (Goldstein and Buhuniak, 2005). Within the survey (Question 5), the researcher will ask the public educators to give their Tennessee Value Assessment composite growth score. This number (1-5) will be used as a dependent variable within the research.

The researcher then cross-referenced the Tennessee Value Assessment composite growth score of each individual teacher to determine if there is a correlation between teacher perception of grading policies and their composite academic growth. In previous research, the researcher has found a correlation between teachers' perceptions of grading practices and their effectiveness at the school level.

Another purpose of this research was to determine if there was a significant difference in the perception of assessment and grading practices between teachers that have had formal assessment training and those that have not had training. The researcher gained insight into this topic by asking survey participants what classes on assessment had they taken throughout their college curriculum in question 4 of the survey.

The researcher also set out to ascertain if there was a significant difference in the perception of assessment and grading practices between teachers that are more experienced and teachers that are less experienced. Question 3 of the survey help glean the information needed to determine if there was a significant difference in the perception of grading practices between teachers with more/less experience.

Data Collections and Consent Process

The researcher began this process by obtaining approval from the University of Memphis' Internal Review Board (IRB). This required forms that were submitted and approved by the IRB board. Once the board granted permission, the researcher began contacting superintendents from across West Tennessee.

The researcher contacted 11 superintendents from West Tennessee to gain authorization to survey teachers. The researcher called the superintendents to explain the specifics of the study and the survey. There were many superintendents opted not to allow their teachers to take part in this survey/research. The most common reason for declining to participate was the concern that teachers were asked to provide their Tennessee Value-Added score. This score reflects the effectiveness of teachers and is highly confidential.

The researcher explained to superintendents that this 26-question survey was administered via Survey Monkey. Survey Monkey offers an option that allows respondents' IP addresses to remain confidential to everyone, including the researcher. The survey results would remain confidential and if there are any identifiable marks, they will be removed after the data are analyzed. Finally, 3 superintendents agreed to allow research to be allowed within their districts.

The researcher sent an email to the superintendent/designee containing a brief message attempting to recruit participants from within these districts to take part in the study. The email, along with a link to the survey, was sent to the superintendents. The superintendent or his/her designee forwarded the e-mail, which contained the brief message and the link to the survey, to the teachers of their districts. The following is the

message that was sent by the researcher to the superintendents:

Hello fellow educator. My name is Jeremy Hinson. I am a middle school assistant principal and a doctoral student at the University of Memphis. For my research, I need to better understand your perspective of the grading process. I believe that the insight that teachers bring to this subject is invaluable. No one else has a better understanding of why your school is successful. Your participation is needed for this research to occur. Please take a moment to fill out this 26-question survey. Thank you for your time and consideration. The survey results of your school will remain confidential and all coding and identifiable marks will be removed after the data is analyzed.

https://www.surveymonkey.com/s/DC5Y9NL

This study uses teacher survey responses from three public school districts in northwest Tennessee to assess the influence that teacher perception of grading practice has on their effectiveness. Approximately 240 teachers were surveyed concerning their perceptions of prevailing grading practices. The teacher that took part in this survey received an email from their superintendent/designee containing a link for this study. Those teachers completed the survey having no contact with the researcher.

Data Analysis

Data identifying teachers' perceptions of the prevailing grading practice have been securely collected, studied, and analyzed. The data has been analyzed using Pearson's Chi-Square Goodness-of-Fit test. The Pearson's Chi-Square Goodness-of-Fit test compares an anticipated frequency to an actual frequency. In order to calculate chi-square goodness-of-fit, the researcher must develop the null hypothesis in which no

statistical significance exists. The researcher must also determine a significance level, in this case 0.5, to determine if there are any data that render statistically significant results for an alternate hypothesis.

Summary

This research examined teachers' perceptions of their classroom grading practices. The researcher is interested in analyzing the relationships between effective and less effective teachers' thoughts about classroom assessment. Comparisons were made on teachers' characteristics (state assessed effectiveness, years of experience, and training).

Chapter 4

Results

Introduction

The literature on teacher grading and assessment reveals a century-long acknowledgment of inconsistency within the grading process. Although scholars acknowledge this inconsistency, teachers' subjective judgments continue to be a large factor when determining students' grades. In this chapter are reported the results of this researcher's exploration of teachers' current perceptions of grading and assessment processes and his analysis of what differences in such perceptions may be observed given external assessments of teaching effectiveness, teachers' years of experience, and teachers' level of formal assessment training. Specific research questions that flow from these concerns are as follows:

1. Are teachers equally likely to respond positively to 21 statements concerning grading and assessment practices?

2. Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?

3. Is there a significant difference in the perception of assessment and grading practices between teachers who have had formal assessment training and those who have not had training?

4. Is there a significant difference in the perception of assessment and grading practices between teachers who are more experienced and teachers who are less experienced?

After a brief description of the sample, findings relevant to each of the aforementioned questions will be presented in turn. The chapter will conclude with a short summary of the relevant findings.

Sample

Presented in Table 2 are the demographic characteristics of the 248 individuals who responded to the questionnaire, all of whom were K-12 educators from three West Tennessee school districts. As indicated in the table, there were significantly more female (87.5%) than male (12.5%) respondents who took part. While the imbalance in participants by gender is noteworthy, it is nevertheless unsurprising in light of outcomes reported by the National Center for Education Statistics (NCES). In the 2011-12 Schools and Staffing Survey (SASS), the NCES reported that females made up 79.4% of teachers in Tennessee (U.S. Department of Education, 2012): a percentage approximating those observed for this study.

Also consistent with the SASS survey results are those observed for the percentages of respondents by age. As shown in the accompanying table, almost two-thirds of the educators surveyed for this study indicated that they were between the ages of 30 and 50: an outcome resonating with the findings of the SASS which placed the average age of Tennessee educators at 42.2. years. An additional point of interest in these results is the fact that 33.2% of the educators surveyed were within the first 10 years of their teaching careers. This would suggest that many of these educators in the range of 30 to 50 years are completing their first decade of teaching service.

However, one of the largest imbalances in this particular dataset concerned their TVASS-based effectiveness scores. The Tennessee Value-Added Accountability System

(TVAAS) was used in this study as a measure of academic success for teachers. Developed by Dr. William Sanders, TVASS-based evaluation scores employ data derived from yearly achievement tests given for students in grades 3-8 in five subject areas: math, reading, language arts, social studies, and science (Goldstein, & Buhuniak, 2005). Given these outcomes teachers may receive a growth score that ranges from a low of 1 to a high of 5. In this study, over half of these respondents reported having earned a level 5 (57.9%), with the 42.1% of respondents earning either a level 4 (29.6%) or lower (12.5%).

To enable the analyses proposed for Research Questions 2 through 4 groups based on the aforementioned demographic variables were kept large and relatively equal in size. Hence, by years of experience, those with 15 years or fewer constituted one analytic grouping (52.6%), while those with more than 15 constituted another (47.4%). Similarly, by their TVASS scores, teachers indicating their having a perfect rating (57.9%) were contrasted with those having a less than perfect rating (42.1%).

Table 2

Demographic Characteristics of Respondents (N = 248)

Characteristic		*f*	*%*
Gender			
	Male	31	12.5
	Female	217	87.5
Age			
	20-29	29	11.8
	30-39	80	32.7
	40-49	83	33.9
	Over 49	53	21.6
Experience			
	1-10 years	82	33.2
	11-15 years	48	19.4
	16-20 years	38	15.4
	20-25 years	43	17.4
	Over 25 years	36	14.6
Coursework			
	Yes	155	62.5
	No	93	37.5
TVASS Score			
	1	1	.4
	2	6	2.5
	3	23	9.6
	4	71	29.6
	5	139	57.9

Research Question 1

Shown in Table 3 are the results pertinent to the first research question that asks "whether teachers are equally likely to respond positively to 21 statements concerning grading and assessment practices." In arriving at an answer to this question, a variation on the chi-square test was employed that checks for the "goodness of fit" between an observed pattern of response frequencies against a hypothesized pattern of response frequencies. In this case, the hypothesized pattern was that the percentages of the respondents agreeing and disagreeing with each of the 21 statements pertaining to grading and assessment practices would approximate a 50/50 distribution. Indicative of the degree to which the observed distribution departs from the hypothesized distribution is the effect size, "w." The larger the value of "w", the further the frequencies are from being equal (50/50 split).

Across the survey's 21 items, responses to four did not appear to be statistically significantly different from the hypothesized 50/50 split. For the following four items—all of which are marked with an "ns," respondents were just as likely to respond affirmatively as negatively:

- Item 5 - Grades are based on the degree to which students participate in class. (55.3% Non-agreement, $w = 0.11$)

- Item 9 - Grades are based on students' ability to turn assignments in on time. (54.1% Non-agreement, $w = 0.08$)

- Item 14 - If a student fails to complete an assignment, I will subtract grade points progressively until the assignment is turned in. (52.5% Non-agreement, $w = 0.05$)

- Item 15 - I often give students opportunities to earn extra credit. (42.1% Non-

agreement, $w = 0.16$)

Absent these items, those remaining evidenced a significant departure from a response rate where approximately 50% of the respondents agreed with the item and 50% of the respondents. As indicated by the robust effect size linked to chi square test results, the items evidencing the greatest departure tended to cluster around two sets of issues: 1) the use of questionable criteria for assigning grades and 2) grading as part of one's instructional role.

With respect to items concerning questionable criteria, participants in very large numbers tended to disagree with basing grades on students' in-class behavior (81.0% disagreement, $w = 0.62$) and their completion of homework (79.3% disagreement, $w = 0.59$). On the other hand, participants in similar numbers tended to agree with the questionable practice of making student effort a consideration in judging the value of their work (79.8% agreement, $w = 0.60$).

As regards the practice of grading itself, most of the participants tended not to agree that grading was "the easiest part of their jobs" behavior (87.3% disagreement, $w = 0.75$) or that it was easy to "assess students' achievement with a single grade or score" behavior (89.0% disagreement, $w = 0.78$). At the same time, over half of the respondents also tended to agree that "did not need grades to teach well" as some sort of index of their instructional effectiveness (80.7% disagreement, $w = 0.61$)

Table 3

Item Frequencies and Percentages Compared

Item	Non Agreement		Agreement		χ^2 (1)	*w*
	n	*%*	*n*	*%*		
1. My students' grades are a reflection of my instructional effectiveness.	83	33.7	163	66.3	26.0	0.33
2. I consider student effort when I grade.	50	20.2	197	79.8	87.5	0.60
3. I will pass a failing student if he or she puts forth effort.	158	63.7	90	36.3	18.6	0.27
4. Grades are based on students' completion of homework.	201	81.0	47	19.0	95.6	0.62
5. Grades are based on the degree to which students participate in class. *(ns)*	136	55.3	110	44.7	2.7	0.11
6. Grades are based on a student's improvement.	98	39.7	149	60.3	10.5	0.21
7. Grades are based on students' attendance.	156	63.4	90	36.6	17.7	0.27
8. I consider student ability in grading.	75	30.5	171	69.5	37.5	0.39
9. Grades are based on students' ability to turn assignments in on time. *(ns)*	133	54.1	113	45.9	1.6	0.08
10. Grades are based on students' behavior in class.	195	79.3	51	20.7	84.3	0.59
11. Grades are based on students' ability to follow directions.	84	34.0	163	66.0	25.3	0.32
12. If a student fails a test, I will offer him/her a second chance to take the test.	93	38.0	152	62.0	14.2	0.24

(Table continues)

Table 3 (continued)

Item	Non Agreement		Agreement		$\chi^2 (1)$	w
	n	*%*	*n*	*%*		
13. If a student fails to complete an assignment, I will assign him/her a grade of zero.	178	72.7	67	27.3	50.3	0.45
14. If a student fails to complete an assignment, I will subtract grade points progressively until the assignment is turned in. *(ns)*	128	52.5	116	47.5	0.6	0.05
15. I often give students opportunities to earn extra credit. *(ns)*	104	42.1	143	57.9	6.2	0.16
16. I often confer with my colleagues on grading criteria.	74	30.1	172	69.9	39.0	0.40
17. Grading is the easiest part of my role as a teacher.	213	87.3	31	12.7	135.8	0.75
18. It is easy for me to assess student achievement with a single grade or score.	218	89.0	27	11.0	148.9	0.78
19. It is difficult to measure student effort.	166	67.2	81	32.8	29.3	0.34
20. I need grades to teach well.	197	80.7	47	19.3	92.2	0.61
21. Grading can help me improve instruction	95	38.5	152	61.5	13.2	0.23

Research Question 2

This question concerns the extent to which teacher effectiveness interacts with or moderates teachers' perceptions of grading and assessment practices asking "Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?" To this end, crosstabs of teacher responses to each item were constructed by their TVASS-based effectiveness ratings as being equal to Level 5 or less than Level 5. Based on the comparison of the observed cell frequencies to the expected cell frequencies, chi-square tests of independence were conducted and phi coefficients denoting the strength of the relationship were computed for each of the 21 items.

Of these 21 tests, only two were statistically significant but two others that approached significance also seem to warrant attention. At a ratio of nearly 80% to 50%, more often did very highly effective teachers (76.8%) than their less effective counterparts (50%) tend to agree that students' grades were a reflection of their own instructional effectiveness ($\phi = 0.28$, $p < .001$). However, the reverse was true when teachers were asked about considerations of student ability in assigning grades: where some 80% of less effective teachers indicated that they agreed with the practice, only some 62.3% of the more effective teachers did so ($\phi = -0.19$, $p = .003$).

In addition to these two items are two others that only approach statistical significance but yield effect sizes that merit consideration. While most of the teachers surveyed disagreed with the practice of assigning no points for missed work, some 77.4% of effective teachers indicated that they refrained from the practice, compared to only 66% of less effective teachers ($\phi = -0.13$, $p = .053$). Interestingly and perhaps

51

Table 4

Item Frequencies and Percentages Compared by Teachers' TVASS-Based Evaluation

Scores

Item	Evaluation ≤ 5		Evaluation $= 5$		ϕ	p
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
1. Reflects effectiveness.	50.0 (50)	50.0 (50)	23.2 (32)	76.8 (106)	0.28	0.000
2. Effort considered.	18.8 (19)	81.2 (82)	21.6 (30)	78.4 (109)	-0.03	0.599
3. Effort mitigates failure.	70.3 (71)	29.7 (30)	59.7 (83)	40.3 (56)	0.11	0.091
4. Homework considered.	78.2 (79)	21.8 (22)	82.0 (114)	18.0 (25)	-0.05	0.464
5. Class participation considered.	52.0 (52)	48.0 (48)	59.4 (82)	40.6 (56)	-0.07	0.255
6. Improvement considered.	38.6 (39)	61.4 (62)	42.8 (59)	57.2 (79)	-0.04	0.520
7. Attendance considered.	63.0 (63)	37.0 (37)	64.0 (89)	36.0 (50)	-0.01	0.870
8. Ability considered.	20.0 (20)	80.0 (80)	37.7 (52)	62.3 (86)	-0.19	0.003
9. Met deadlines considered.	51.0 (51)	49.0 (49)	56.5 (78)	43.5 (60)	-0.05	0.399
10. Behavior considered.	76.2 (77)	23.8 (24)	81.2 (112)	18.8 (26)	-0.06	0.355

Table 4 (continued)

Item	Evaluation < 5		Evaluation = 5		φ	p
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
11. Requirements considered.	35.6 (36)	64.4 (65)	34.1 (47)	65.9 (91)	0.02	0.799
12. Second chances provided	40.0 (40)	60.0 (60)	35.8 (49)	64.2 (88)	0.04	0.506
13. No points for missed work.	66.0 (66)	34.0 (34)	77.4 (106)	22.6 (31)	-0.13	0.053
14. Late work reduces points.	54.5 (54)	45.5 (45)	51.1 (70)	48.9 (67)	0.03	0.600
15. Extra credit given.	43.6 (44)	56.4 (57)	39.9 (55)	60.1 (83)	0.04	0.565
16. Colleagues consulted..	32.0 (32)	68.0 (68)	29.7 (41)	70.3 (97)	0.02	0.705
17. Grading is easy.	89.9 (89)	10.1 (10)	85.4 (117)	14.6 (20)	0.07	0.306
18. Single grade/score sufficient.	88.0 (88)	12.0 (12)	89.9 (124)	10.1 (14)	-0.03	0.651
19. Effort difficult to measure.	58.4 (59)	41.6 (42)	73.2 (101)	26.8 (37)	-0.16	0.016
20. Grades needed to teach well.	78.0 (78)	22.0 (22)	81.8 (112)	18.2 (25)	-0.05	0.474
21. Grading improves instruction.	39.6 (40)	60.4 (61)	37.0 (51)	63.0 (87)	0.03	0.677

counterintuitively, while the majority the sampled teachers indicated their belief that "effort was NOT especially difficult to measure," over 70% of teachers in the highly effective group did so (73.2%) compared to about 60% of teachers in the less effective group (58.4%). In this instance, teachers with lower TVASS-based evaluation scores seemed somewhat more troubled at the level of subjectivity at making such judgements than those with higher TVASS-based scores ($\phi = -0.16$, $p = .016$).

Research Question 3

This question addresses the extent to which coursework in assessment interacts with or moderates teachers' perceptions of grading and assessment practices asking "Is there a significant difference in the perception of assessment and grading practices between teachers who have had formal assessment training and those that have not had training? To this end, crosstabs of teacher responses to each item were constructed by their yes or no responses to a single question about their having had a course in assessment. Based on the comparison of the observed cell frequencies to the expected cell frequencies, chi-square tests of independence were conducted and phi coefficients denoting the strength of the relationship computed were for each of the 21 items.

Perhaps because taking at least one assessment course is now an established part of the teacher education curriculum, that characteristic would seem to discriminates less among present survey respondents than when the instrument was first developed. As a result, none of the items appeared to discriminate across groups, with those having had an assessment course as likely as not to agree with an item as those without formal coursework. Had the respondents been asked for the number of courses or whether the

course was graduate-level or undergraduate-level, more nuanced results might have emerged.

Research Question 4

This question explores the extent to which experience impacts teachers' perceptions of grading and assessment practices asking "Is there a significant difference in the perception of assessment and grading practices between teachers that are more experienced and teachers that are less experienced?" To this end, crosstabs of teacher responses to each item were constructed by their having indicated that they had 15 or fewer or more than 15 years of teaching experience. Based on the comparison of the observed cell frequencies to the expected cell frequencies, chi-square tests of independence were conducted and phi coefficients denoting the strength of the relationship were computed for each of the 21 items.

By years of experience, only one item appeared to distinguish between groups at statistically significant levels. As a factor in assigning grades, nearly three-fourths of less experienced teachers (74.4%) indicated their willingness to consider students' ability to follow directions or, more broadly, meeting the requirements of an assignment, compared to a little more than half of more experienced teachers (57.3%). While the reasons for this relationship remain to be explained, that there is such a relationship is suggested by a robust effect size ($\phi = -0.13$, $p = .053$). Two other items with which less experienced teachers are more likely to agree than their more experienced counterparts concern reducing points for late work ($\phi = -0.11$, $p = .076$) and finding it easy to sum up a student's achievement with a single grade or score ($\phi = -0.12$, $p = .053$). With respect to the aforementioned grading practice, over 50% of the less experienced group were likely

Table 5

Item Frequencies and Percentages Compared by Teachers' Taking or Not Taking Formal

Coursework in Assessment

Item	Had Course		No Course		φ	*p*
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
1. Reflects effectiveness.	32.3 (50)	67.7 (105)	36.3 (33)	63.7 (58)	-0.04	0.521
2. Effort considered.	22.1 (34)	77.9 (120)	17.2 (16)	82.8 (77)	0.06	0.356
3. Effort mitigates failure.	61.9 (96)	38.1 (59)	66.7 (62)	33.3 (31)	-0.05	0.453
4. Homework considered.	83.2 (129)	16.8 (26)	77.4 (72)	22.6 (21)	0.07	0.259
5. Class participation considered.	58.1 (90)	41.9 (65)	50.5 (46)	49.5 (45)	0.07	0.252
6. Improvement considered.	40.0 (62)	60.0 (93)	39.1 (36)	60.9 (56)	0.01	0.893
7. Attendance considered.	62.3 (96)	37.7 (58)	65.2 (60)	34.8 (32)	-0.03	0.650
8. Ability considered.	28.6 (44)	71.4 (110)	33.7 (31)	66.3 (61)	-0.05	0.398
9. Met deadlines considered.	52.3 (80)	47.7 (73)	57.0 (53)	43.0 (40)	-0.05	0.473
10. Behavior considered.	81.0 (124)	19.0 (29)	76.3 (71)	23.7 (22)	0.06	0.378

(Table continues)

Table 5 (continued)

Item	Had Course		No Course		φ	p
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
11. Requirements considered.	34.8 (54)	65.2 (101)	32.6 (30)	67.4 (62)	0.02	0.721
12. Second chances provided	38.3 (59)	61.7 (95)	37.4 (34)	62.6 (57)	0.01	0.882
13. No points for missed work.	74.7 (115)	25.3 (39)	69.2 (63)	30.8 (28)	0.06	0.356
14. Late work reduces points.	50.7 (77)	49.3 (75)	55.4 (51)	44.6 (41)	-0.05	0.469
15. Extra credit given.	39.4 (61)	60.6 (94)	46.7 (43)	53.3 (49)	-0.07	0.256
16. Colleagues consulted..	27.9 (43)	72.1 (111)	33.7 (31)	66.3 (61)	-0.06	0.339
17. Grading is easy.	85.6 (131)	14.4 (22)	90.1 (82)	9.9 (9)	-0.07	0.309
18. Single grade/score sufficient.	90.3 (139)	9.7 (15)	86.8 (79)	13.2 (12)	0.05	0.405
19. Effort difficult to measure.	69.0 (107)	31.0 (48)	64.1 (59)	35.9 (33)	0.05	0.428
20. Grades needed to teach well.	80.3 (122)	19.7 (30)	81.5 (75)	18.5 (17)	-0.02	0.809
21. Grading improves instruction.	34.8 (54)	65.2 (101)	44.6 (41)	55.4 (51)	-0.10	0.129

Table 6

Item Frequencies and Percentages Compared by Teachers' Years of Teaching

Experience

Item	15 Years or Fewer		Above 15 Years		ϕ	p
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
1. Reflects effectiveness.	38.0 (49)	62.0 (80)	29.3 (34)	70.7 (82)	0.09	0.152
2. Effort considered.	23.3 (30)	76.7 (99)	17.1 (20)	82.9 (97)	0.08	0.230
3. Effort mitigates failure.	63.8 (83)	36.2 (47)	63.2 (74)	36.8 (43)	0.01	0.922
4. Homework considered.	82.3 (107)	17.7 (23)	79.5 (93)	20.5 (24)	0.04	0.573
5. Class participation considered.	51.6 (66)	48.4 (62)	59.0 (69)	41.0 (48)	-0.07	0.244
6. Improvement considered.	43.4 (56)	56.6 (73)	35.9 (42)	64.1 (75)	0.08	0.229
7. Attendance considered.	64.3 (83)	35.7 (46)	62.1 (72)	37.9 (44)	0.02	0.713
8. Ability considered.	32.6 (42)	67.4 (87)	28.4 (33)	71.6 (83)	0.04	0.486
9. Met deadlines considered.	50.0 (65)	50.0 (65)	58.3 (67)	41.7 (48)	-0.08	0.195
10. Behavior considered.	76.6 (98)	23.4 (30)	82.1 (96)	17.9 (21)	-0.07	0.291

(Table 6 continues)

(Table 6 continued)

Item	15 Years or Fewer		Above 15 Years		φ	p
	No % (n)	Yes % (n)	No % (n)	Yes % (n)		
11. Requirements considered.	25.6 (33)	74.4 (96)	42.7 (50)	57.3 (67)	-0.18	0.004
12. Second chances provided	42.6 (55)	57.4 (74)	33.0 (38)	67.0 (77)	0.10	0.124
13. No points for missed work.	70.3 (90)	29.7 (38)	75.0 (87)	25.0 (29)	-0.05	0.413
14. Late work reduces points.	47.2 (60)	52.8 (67)	58.6 (68)	41.4 (48)	-0.11	0.076
15. Extra credit given.	44.2 (57)	55.8 (72)	40.2 (47)	59.8 (70)	0.04	0.524
16. Colleagues consulted..	28.9 (37)	71.1 (91)	30.8 (36)	69.2 (81)	-0.02	0.750
17. Grading is easy.	86.6 (110)	13.4 (17)	87.9 (102)	12.1 (14)	-0.02	0.759
18. Single grade/score sufficient.	85.3 (110)	14.7 (19)	93.0 (107)	7.0 (8)	-0.12	0.053
19. Effort difficult to measure.	65.1 (84)	34.9 (45)	69.2 (81)	30.8 (36)	-0.04	0.493
20. Grades needed to teach well.	78.1 (100)	21.9 (28)	83.5 (96)	16.5 (19)	-0.07	0.291
21. Grading improves instruction.	38.0 (49)	62.0 (80)	39.3 (46)	60.7 (71)	-0.01	0.830

to engage (52.8%), compared to about 40% of the more experienced group (41.4%). With respect to the aforementioned belief about grading, about 15% of the less experienced group were likely to agree (14.7%), compared to about half that percentage of the more experienced group (7.0%).

Summary

Consistent with the literature, teachers do not always evoke valid criteria when trying to assess student achievement. As indicated by teachers' responses to many items, factors like a student's effort, ability, willingness to comply are likely to creep into their assessment and grading practices, making their report of student academic performance as summarized in a grade something not purely and precisely that. While such factors as formal training and years of experience may moderate some aspects of teacher grading and assessment practices, they as yet do not appear to have a systemic influence on what teachers believe about grading and assessment and how they discharge those instructional duties.

Chapter 5

Discussion

Introduction

The literature on teacher grading and assessment reveals a century-long acknowledgment of inconsistency within the grading process. Although scholars acknowledge this inconsistency, teachers' subjective judgments continue to be a large factor when determining students' grades. Scholars' viewpoints have been evident; yet practitioners have not voiced an opinion as to why they continue to utilize the prevailing grading system. Duncan and Noonan (2007) contended that there is insufficient research on the perspective of practitioners concerning grading. The overriding purpose of this study was to gain insight into teachers' perspectives on grading.

The researcher gained this insight by determining the degree to which state-assessed effectiveness, years of experience, and assessment training have on teachers' perception of grading practices. The researcher was also interested in measuring the level of agreement among teachers regarding 21 questions concerning grading practices. In this chapter the researcher reviews the methodology, presents answers to the research questions, and discusses those findings.

Review of the Methodology

This research examined teachers' perceptions of their classroom grading practices. The researcher was interested in analyzing the relationships between effective teachers' viewpoints on classroom assessment. Comparisons were made on teachers related to factors such as teacher effectiveness, training, and years of experience.

Approximately 240 teachers were surveyed concerning their perceptions of prevailing grading practices.

Data identifying teachers' perceptions of the prevailing grading practice have been securely collected, studied, and analyzed. The data has been analyzed using Pearson's Chi-Square Goodness-of-Fit test. The Pearson's Chi-Square Goodness-of-Fit test compares an anticipated frequency to an actual frequency. In order to calculate chi-square goodness-of-fit, the researcher must develop the null hypothesis in which no statistical significance exists. The researcher must also determine a significance level, in this case 0.5, to determine if there are any data that render statistically significant results for an alternate hypothesis.

Discussion of Results

The researcher determined multiple differences in the perception of assessment and grading practices among teachers for some of the factors studied. The following research questions directed this study.

1. Are teachers equally likely to respond positively to 21 statements concerning grading and assessment practices?

2. Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?

3. Is there a significant difference in the perception of assessment and grading practices between teachers who have had formal assessment training and those that have not had training?

4. Is there a significant difference in the perception of assessment and grading practices between teachers who are more experienced and teachers who are less experienced?

Research Question 1

Are teachers equally likely to respond positively to 21 statements concerning grading and assessment practices?

In analyzing the responses to question 1, the outright answer is yes for four of the items. However, the data revealed a significant difference for 10 of the items. Four items responses were equally divided. Five of the items teachers responded negatively; and for one item teachers responded positively. Responses to the other 14 items were not statistically significant.

Equally Divided Items.

Responses to the following four questions revealed that teachers responded equally: [1]

- Item 5 - Grades are based on the degree to which students participate in class. (55.3% Non-agreement, $w = 0.11$)

- Item 9 - Grades are based on students' ability to turn assignments in on time. (54.1% Non-agreement, $w = 0.08$)

- Item 14 - If a student fails to complete an assignment, I will subtract grade points progressively until the assignment is turned in. (52.5% Non-agreement, $w = 0.05$)

- Item 15 - I often give students opportunities to earn extra credit. (42.1% Non-agreement, $w = 0.16$)

Disagreement

When responding to the questions regarding homework and student behavior teachers tended to disagree. They disagreed by 79.3% and 81.0% respectfully:

- Item 4 – Grades are based on students' completion of homework. (79.3% Non-agreement, w = 0.59)

- Item 10 – Grades are based on students' behavior in class. (81.0% Non-agreement, w = 0.62)

Agreement

With regard to effort, teachers tended to agree at 79.8%. Teachers agreed with the following item:

- Item 3 – I will pass a failing student if he or she puts forth effort. (79.8% Agreement, w = .60)

Hodgepodge Grading

Teachers responded to 7 of the 24 items in a different manner. They responded equally to 4 items, they disagreed with 2 items, and they agreed with one item. Based on the literature, one would tend to think that they would disagree with all 7 items. All of these items listed above (Items 3, 4, 5, 9, 10, 14, and 15) are examples of what Guskey (2006a) calls "hodgepodge grades". Hodgepodge grades include "elements of achievement, attitude, effort, and behavior" (p. 671). However, each teacher utilizes autonomy when collecting evidence of student growth and achievement (Cizek, Fitzgerald, Shawn, & Rachor, 1996). This evidence can be taken from presentations, projects, quizzes, writing tasks, tests, and/or demonstrations (Guskey, 2002b); and this evidence are subjective to teachers' perception of students' ability, attitude, effort, and behavior. Specifically this teacher autonomy leaves a large variation on the final grade

depending on what factors are seen as most important to the teacher. Hodgepodge grading transpires due to teachers that are forced to report evidence from several different sources into one symbol (Guskey, 2006a).

Effort in Grading

Interestingly, the majority of respondents answered in the affirmative to Item 3: I consider student effort when I grade. Therefore this finding acknowledges that teachers' practices are consistent with what researchers have written. Teachers often utilize effort as a component in determining grades. (Brookhart 1994; Cizek et al., 1996; McMillan et al, 2002; Stiggins et al., 1989). On the other hand, researchers acknowledge and continually argue against the practice of considering student effort when grading. In fact, researchers have urged against the use of process criteria (effort) in grading for years. Stiggins and Knight (1997) wrote that in terms of effort, "definitions of trying hard vary greatly from teacher to teacher" (p. 418). O'Connor (2002) said, "strong effort, active participation, and positive attitude are highly valuable attributes, but they are reporting variables, not grading variables" (p. 100). However, this study concludes that these practitioners overwhelmingly adhere to this practice. In addition, teachers often factor effort in the case of underachieving students (Stiggens et al, 1989) because "these students find the relationship between high effort and low grades frustrating and often express their frustration with indifference, deception, or disruption" (Guskey, 2006, p. 672).

All students are not created equal. For example, some students are naturally more academically talented than other students. Therefore, with no true growth measure for

65

classroom teachers outside of state assessed testing, teachers level the playing fielding by factoring the effort of underachieving students.

Practitioners understand the practical usefulness and limitations of each of Brookhart's (1993) grading criteria:

1. Product (Academic achievement measured through exams, papers, projects etc.)

2. Progress (Academic growth measured between two points in time.)

3. Process (Student effort perceived through participation and work habits)

Therefore, the limitations within each of these grading criteria prevent practitioners from using them exclusively. Guskey (2006a) expressed this sentiment:

> Because of concerns about student motivation, self-esteem, and social consequences of grades, few teachers use only product criteria in determining grades. Instead, most routinely base their grading procedures on some combination of all three types of evidence. Many also vary their grading criteria from student to student, taking into account individual circumstances. (p. 672)

Other items that rendered statistically significant results to research question one. The following questions asked teachers about their opinions of the prevailing grading system in the United States. Overwhelmingly, teachers tended to disagree with the following items 87.3%, 89.0%, and 80% respectively. These items of interest are: [1]

Disagreement

- Item 17 – Grading is the easiest part of my role as a teacher. (87.3% Non-agreement, w = 0.75)

- Item 18 – It is easy for me to assess student achievement with a single grade of

---

[1] The symbol "w" is the effect size associated with the Chi-Square "equal frequencies" test. The larger the value the bigger the differences from a 50/50 split.

score. (89.0% Non-agreement, w = 0.78)

- Item 20 – I need grades to teach well. (80.7% Non-agreement, w = 0.61)

Teachers' Voice

Teachers overwhelmingly disagreed with the above items (17, 18, and 20).

Moreover, the majority of teachers in the study believed that the prevailing grading

system was challenging and that reporting student achievement with a single grade was

difficult. Additionally, this is in line with what researchers have proposed. The

subjective nature of grading and reporting creates difficulty for teachers to report

accurately. Therefore, the importance of separating "hodgepodge" grading from

achievement is paramount for the sake of all stakeholders (Jung & Guskey, 2007).

Moreover, the presence of nonacademic factors within academic indicator of proficiency

is neither accurate nor helpful to stakeholders (Brookhart, 2009; Guskey & Bailey, 2001).

Concluding that, nonacademic factors need be reported separately in order to create a

more accurate account of the students' knowledge (Guskey, 2006a); and to streamline the

grading process for teachers.

Many researchers consider the abundance of variables within the learning process

to comprehensively report learning with a single reporting mechanism (Allison &

Friedman, 1995). However, instead of relying on report cards, some researchers

recommend a comprehensive reporting system (Guskey, 2002). Guskey (2002)

envisioned a reporting system including the following items: report cards, narratives

attached to reports cards, test results, phone communication, progress reports, school

open houses focusing on assessment procedures, personal correspondences, portfolios of

student work, exhibits of student work, and conferences (Webster, 2011).

Practitioners also stated in large numbers that they did not need grades in order to teach well. Chapman and Ashbaugh (1925) cited concerns about the validity and reliability of grades. The researchers recommended the elimination of the formal grading system dated back to the 1920s. Additionally, Kohn (1993) alluded to the elimination of grading by expressing concern that grades were harmful motivators for students. Moreover, Brookhart (2009) stated, "in a perfect world there would be no grades – at least not as we know them now" (p. iii).

Research Question 2

Is there a significant difference in the perception of assessment and grading practices between teachers who score Level 5 and those who do not?

The outright answer to research question 2 is yes for three items. The other 21 items did not present a significant statistical difference in teachers' perceptions. The following three items rendered statistically significant results pertaining to research question two:[2]

- Item 1 – My students' grades are a reflection of my instructional effectiveness. ($\phi = 0.28$)

- Item 8 – I consider student ability in grading. ($\phi = -0.19$)

- Item 19 – It is difficult to measure student effort. ($\phi = -0.16$)

Reflects Effectiveness

Specifically, the survey results show that highly effective teachers believe that grades are a reflection of their own instructional effectiveness. Therefore, highly

---

[2] The symbol ϕ denotes the phi coefficient, which is a measure of association for two binary variables.

effective teachers believe they can lead students to success despite the students' background or barriers. They have the mindset that they can cause the positive change needed for the success of their students. Most importantly, the fact that these teachers take ownership of this responsibility makes all the difference. These results are comparable to a study published in 2012 called "The Irreplaceables". "Compared to low-performing teachers, Irreplaceables (highly effective teachers) are more likely to consider that effective teachers can help students overcome out-of-school challenges and are more likely to understand their own effectiveness". (The Irreplaceables, 2012, p. 3)

We as human beings have some innate requirements. After physiological needs, Maslow suggests that humans need to experience a sense of safety followed by love/belonging. Moreover, a sense of belonging can be secured through relationships cultivated by the mutual respect of individuals working toward a common goal. On the other hand, the subjective nature of current grading systems causes an adversarial relationship between teacher and students (Kohn, 1999).

Ability Considered

Less effective teachers tend to agree that they considered student ability when grading. This research shows that 80% of teachers who have value-added scores of 1-4 agreed that they consider student ability when grading. However, only 62% of level 5 teachers agreed with this statement.

Effort Measurement

This research confirms that highly effective (Level 5) teachers perceive that they do not find it difficult to measure student effort. At of rate of 73%, highly effective teachers reported that it is not challenging to determine their students' effort when

grading. On the other hand, less effective teachers (Levels 1-4) acknowledged to a lesser

degree (58.4%) that it was not challenging to determine students' effort when grading.

Research Question 3

Is there a significant difference in the perception of assessment and grading practices

between teachers who have had formal assessment training and those that have not had

training?

    Formal Assessments.

    Results indicated that normal assessment training made no discernible difference

in teachers' perception of grading. The researcher anticipated that there would be a

difference in perception between those practitioners who had gained

education/knowledge on the topic of assessment and those that had not. These results are

interesting because they highlight that higher education curriculum did not have the

impact on teachers' perceptions of grading practices. Additionally, there seems to be a

disconnect between college education programs and educational practitioners. K-12

educators have previously noted the substandard arena of collegiate education programs.

Arnie Duncan, the Secretary of Education, is quoted as saying, "America's university

based teacher preparation programs need revolutionary change". Levine (2006), who is a

noted author and scholar of higher education wrote, "Teacher education is the Dodge City

of the education world. Like the fabled Wild West town, it is unruly and chaotic" (p. 1).

Hattie (2009) wrote:

        I have sat through many meetings where colleagues have decided on the essential

        core knowledge and experiences that should be taught to teacher education

        students. In every place this has been long and often vexed discussion and every

time the "core" knowledge decided on by the group has been different. There is

not set of essential experiences that must be taught, let alone a correct order for

students to become a teacher. Moreover, it seems surprising that the education of

new teachers seems so data-free; maybe this is where future teachers learn how to

ignore evidence, emphasize craft, and look for positive evidence that they are

making a difference. (p. 110)

The fact that formal education in assessment did not impact the perspectives of

the participant of this survey is not surprising. The participants of this study are likely to

have studied at different universities emphasizing different ideologies on assessment

processes and practices.

Research Question 4

Is there a significant difference in the perception of assessment and grading practices

between teachers who are more experienced and teachers who are less experienced?

Three items from research question 4 proved to have statistical significance. The

other 21 items did not result in any statistical difference. The following three items

exhibited a medium [3]effect:

- Item 11 – Grades are based on students' ability to follow directions. ($\phi = -0.18$)

- Item 14 – If a student fails to complete an assignment, I will subtract grade points

  progressively until the assignment is turned in. ($\phi = -0.11$)

- Item 18 – It is easy for me to assess student achievement with a single grade or

  score. ($\phi = -0.12$)

---

[3]
   The symbol ϕ denotes the phi coefficient, which is a measure of association for two binary variables.

Following Directions/Subtracting Points

Teachers with less 15 years of teaching tended to agree, at a higher rate than their more veteran counterparts, that grades are based on students' ability to follow directions. However, teachers with more than 15 years of experience tended to disagree with the practice of subtracting grade points on late assignments. In addition, grades based on following directions (Item 11) and subtracting points for incomplete assignments (Item 14) are examples of teachers using grades to cultivate positive behavior in students.

In a new teacher's manual from 1917, Woofter named five habits that teachers were to promote in every student: promptness, industry, politeness, neatness and order, and self-control. To this day, in an effort to promote positive learning behaviors and behaviors that demonstrate engagement, teachers include these types of behaviors in their grading calculations (Brookhart, 2009; Hendrickson & Gable, 1997, Webster, 2011).

Single Grade

The more experienced group of teachers had a tendency to respond that they it was difficult to report student achievement with a single grade or score. This is another case where practitioner perspective is congruent with researcher's studies. Many researchers believe there are too many variables within the learning process to comprehensively report learning with a single reporting mechanism (Allison & Friedman, 1995).

Summary of Major Findings

Highly Effective Teachers. This study provides recommendations for teaching and learning. Highly effective teachers versus their less effective counterparts provided more discrepancy than any other variables introduced in this research study.

Rubie-Davies, et al (2011) noted that teachers' beliefs guide their actions, and teachers' actions affect student academic growth and achievement. McMillian (2007) argued that research on teacher beliefs should expand. Therefore, comprehending how teacher beliefs impact teaching practices, favorably or negatively, could result in a breakthrough in student learning outcomes (Koloi-Keaikitse, 2012).

There are no distinguishing traits that encapsulate highly effective teachers. Highly effective teachers are represented in different experience levels, educations levels, and teaching styles. However, highly effective teachers' actions seem to be guided by their beliefs. Therefore these beliefs and ensuing actions set them apart. This study shows that highly effective teachers believe that students' grades are a reflection of the teachers' own instructional effectiveness, which is a belief that can be cultivated in all teachers. Highly effective teachers also do not consider students' ability as much as their non effective counterparts when grading.

Formal Assessments. Formal assessment training made no discernible difference in teachers' perception of grading. Those that had gained education on the topic of assessment did not exhibit a perception shift from those that had not received assessment training. These results are interesting and highlight that higher education curriculum might not have the impact that is needed to produce effective educators.

Effort in Grading. The majority of participants (79.8%) surveyed tended to agree with the fact that they considered student effort when grading. Researchers have argued that this is a questionable practice by citing the fact that the perception of student effort is difficult to measure and varies from teacher to teacher. Highly effective teachers report that they can gauge student effort at a higher rate than their less effective counterparts.

Recommendations for Educators/Additional Research

This study can serve as complementary research for educators. Future researchers could use a larger sample size to study teacher perception. Researchers could test these findings at a state or even national level. If a change in the prevailing grading practice is going to occur, we need to obtain teacher input on a large scale.

Today's educational structures are deeply rooted in measurement. We obtain evidence of student growth through school and district data, state achievement through public report cards, and even examine national exam data such as NAEP to determine the degree to which our student are learning.

As previously stated in this paper, educators within the state of Tennessee receive a growth score that ranges from a low of 1 to a high of 5. We exact teacher effectiveness based on how much growth their students collectively exhibit within the school year. Why are we grading students on a different scale than we grade teachers? Most teachers are not as concerned about grades as they are about their students' growth scores. The growth score is a true representation of how well teachers convey information to students. If we have the capacity to calculate student growth (value-added systems), why do we continue to use the subjective practice of grading?

The researcher contends that formative standards-based assessments should be developed and given to students twice every nine-week period. These assessments would be objective and consistent providing all stakeholders with a true representation of student growth and achievement.

Conclusion

All of the social norms that we abide by, outside of those dictated by physiological needs, are models created by human beings. These models are ingrained so deeply that we often adhere to these norms collectively without giving them much thought. The grading practice predominantly used throughout the United States is a good example of an ingrained social norm.

Those that have examined the status quo acknowledge the shortcomings of the prevailing grading system. This research study shows that teachers often agree with practitioners about the specific shortcomings of the grading process. However, grading practices have not evolved over the last century because practitioners are bound by the parameters of the reporting system that is currently in place. We must now look to educational leaders and policy makers to develop the change needed in the prevailing grading system.

References

Abbott, S. (Ed.). (2014, August). Hidden curriculum. In The glossary of education reform. Retrieved from http://edglossary.org/hidden-curriculum

Academic Benchmarks. (2015). *Common Core State Standards Adoption Map.* Retrieved from http://academicbenchmarks.com/common-core-state-adoption-map on June 22, 2016.

Agnew, E. J. (1985, March-April). *The grading policies and practices of high school teachers.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Airasian, P. W. (2005). *Classroom assessment: Concepts and applications* (5th ed.). New York, NY: McGraw-Hill Higher Education.

Allen, J. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.

Allison, E., & Friedman, S. (1995). Reforming report cards. *Executive Educator, 17*(1), 38-39.

Ayiro L. P. (2012). Analysing and Interpreting Data. In *A functional approach to educational research methods and statistics: qualitative, quantitative, and mixed methods approaches*. Lewiston, NY: Edwin Mellen Press.

Baker, C. (2013, May 24). Common Core State Standards attract controversy across the U.S. *Deseret News National*. Retrieved from national.deseretnews.com/article/343/Common-Core-State-Standards-attract-controversy-across-US.html?pg=all

Banathy B. H. (1996). *Designing social systems in a changing world*. New York, NY: Plenum.

Bertalanffy, L. V. (1956). General systems theory. *General Systems*, *I*(1), 1-10.

Bidwell, A. (2014, March). A guide to Common Core. *U.S. News & World Report.* Retrieved from www.usnews.com/news/special-reports/a-guide-to-common-core/articles/2014/03/06/the-politics-of-common-core

Bigham Baron, P. A. (2000). *Consequential validity for high school grades: What is the meaning of grades for senders and receivers?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80 (2)*, 139-148. Retrieved from

http://www.spd.dcu.ie/site/teaching_today/documents/Raisingstandardsthroughclassroomassessment.pdf

Bliem, C.L., & Davinroy, K.H. (1997). Teachers' beliefs about assessment and instruction in literacy (CSE Technical Report 421). Los Angeles, CA: National Center for Research on Evaluation, Standards & Student Testing

Blount, H. P. (1997). The keepers of numbers: Teachers' perspectives on grades. *The Educational Forum, 61*(4), 329-334.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30 (2), 123-142.*

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

Brookhart, S. M. (2009). Grading (2nd Ed.). Upper Saddle River, NJ: Pearson, Merrill, Prentice- Hall.

Burstein, L. (1978). Secondary analysis: An important resource for educational research and Evaluation. *Educational Researcher, 7*(5), 9-12.

Cantu, D. A. (2001). *An investigation of the relationship between social studies teachers' beliefs and practice*. Lampeter, Wales: The Edwin Mellen Press.

Chapman, H.B., & Ashbaugh, E.L. (1925). Report cards in American cities. *Educational Research Bulletin, 4*, 289-310.

Cizek, G.J., Fitzgerald, S.M., & Rachor, R. E. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159- 179.

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

Cohen, S. (1974). *Education in the United States: A documentary history* (Vol. 4). New York, NY: Random House.

Common Core State Standards Initiative. (nd). Preparing America's students for college and career. Retrieved from http://www.corestandards.org/about-the-standards/frequently-asked-questions/ on September 1, 2015.

Covington, M. A. (2004). *What should grades mean?* Retrieved from http://www.covingtoninnovations.com/mc/grading.html on July 13, 2016.

Creswell, J. W. (2008). Educational research: Planning, conducting, and evaluating

quantitative and qualitative research. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009) *Internet, mail and mixed-made surveys. The tailored design method* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Dressel, P. (1983). Grades: One more tilt at the windmill. In A.W. Chickering (Ed.), *Bulletin* (p.12). Memphis, TN: Memphis State University.

Duncan, C. R., & Noonan, B. (2007, Spring). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53*(1), 1-21. Retrieved from http://www.education.ualberta.ca/educ/journals/ajer/ 53%20files/53(1)files/A1.html

Durm, M. W. (1993, Spring). An A is not an A is not an A: A history of grading. The Educational Forum, 57. Retrieved from http://www.indiana.edu/~educy520/sec6342/week_07/durm93.pdf January 22, 2014.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Ferguson, R. F. (2009). Achievement Gap [Video]. Retrieved from http://www.learningace.com/doc/2116890/5d2487e763365a45e3217c08ba09ee6c/ achievement-gap-dr-ronald-f-ferguson# on June 10, 2013.

Finkelstein, I.E. (1913). The Marking System in Theory and Practice. Baltimore, MD. Retrieved from https://archive.org/details/markingsystemin02finkgoog on June 4, 2016.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993, Fall). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practices, 12*(3), 23-30.

Fritzberg, G. (2001, May). No shortcuts to excellence. *Sojourners Magazine*, *30*(3), 17. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA74799780&v=2.1&u=tel_a_uof mem&it=r&p=AONE&sw=w

Fritzberg, G. J. (2004). Revise and resubmit: A critical response to Title One of the No Child Left Behind Act. *Journal of Education*, *184*(1), 69-85. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA119203640&v=2.1&u=tel_a_uof mem&it=r&p=AONE&sw=w

Goldstein, J., & Behuniak, P. (2005). Growth models in action: Selected case studies.

Practical Assessment, Research, and Evaluation 10 (11), p.1-17. Retrieved April 4, 2012 from http://pareonline.net/pdf/v10n11.pdf.

Good, W. (1937). Should grades be abolished? *Education Digest, 2*(4), *7-9.*

Goodlad, J.I. (2002). Kudzu, rabbits, and school reform. *Phi Delta Kappan*, *84*(1), 16-23.

Green, R. (2010). *The four dimensions of principal leadership: A framework for leading 21st century schools.* Boston, MA: Pearson.

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

Guskey, T.R. (1996). Reporting on student learning: Lessons from the past—prescriptions for the Future. *Association for Supervision and Curriculum Development, 13*(1), 20-32.

Guskey, T. R. (1997). *Implementing mastery learning* (2nd ed.). Belmont, CA: Wadsworth.

Guskey, T. R. (2000, December). Grading policies that work against standards and how to fix them. *NASSP Bulletin*, *84*(620), 20-29. doi:10.1177/ 019263650008462003

Guskey, T. R. (2001). The backward approach. Journal of Staff Development, 22(3), 60.

Guskey, T. R. (2002a). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan, 83*(10), 775-780. Retrieved from http://www.kappanmagazine.org

Guskey, T. R. (2002b). *Grading and reporting student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (2006a, April). *"It wasn't fair!" Educators' recollections of their experiences as students with grading*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Guskey, T. R. (2006b). Making high school grades meaningful. Phi Delta Kappan, 87(9), 670-675. Retrieved from http://www.allen.kyschools.us/Downloads/Guskey%20high%20school%20gradin g.pdf on July 13, 2016.

Guskey, T.R. (2011). Five Obstacles to Grading Reform, Educational Leadership, Vol. 69, No. 3, November 2011, pp. 16-21. Retrieved from http://www.ascd.org/publications/educational-leadership/nov11/vol69/num03/Five-Obstacles-to-Grading-Reform.aspx June 10, 2016.

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

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

Gutek, G.L. (1986). *Education in the United States: An historical perspective.* Englewood Cliffs, NJ: Prentice Hall.

Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice items writing guidelines for classroom assessment. *Applied Measurement in Education, 15 (3), 309-334.*

Hattie, J. (2009). Visible Learning: a synthesis of over 800 meta-analysis relating to achievement. Routledge, Milton Park, Oxon: UK.

Hattie, J., & Timperley, H. (2007, March). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi:10.3102/003465430298487

Heck, A.O. (1938). Contributions of research to the classification, promotion, marking and certification of pupils. In G. M. Whipple (Ed.), The scientific movement in education (pp. 187-199). Bloomington, IL: Public School.

Heifetz, R.A., Grahow, A., & Linsky, M. (2009). *The practice of adaptive leadership: Tools and tactics for changing your organization and the world.* Cambridge, MA: Harvard Business Press.

Hendrickson, J., & Gable, R. (1997). Collaborative assessment of students with diverse needs: Equitable, accountable, and effective grading. *Preventing School Failure, 41*(4), 159-164.

Johnson, J.K. (2001). *The grading of elementary student performance on a standards-based report card.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3013976)

Katz, D., & Kahn, R. (1966). *The social psychology of organizations.* New York, NY: John Wiley & Sons.

Kennedy, K., Peters, M., & Thomas, M. (2012). *How to use value-added analysis to improve student learning: A field guide for school and district leaders.* Thousand Oaks, CA: Corwin A Sage Company

Kirschenbaum, H., S. B. Simon, & R.W. Napier. (1971). *Wad-ja-get? The grading game in American education.* New York, NY: Hart.

Kliebard, H. M. (2004). *The struggle for the American curriculum: 1893-1958* (2nd ed.). New York, NY: Taylor & Francis Books.

Kohn, A. (1993). *Punished by rewards.* Boston, MA: Houghton Mifflin.

Kohn, A. (1999). From degrading to de-grading. *High School Magazine*, 6(5), 38-43.

Koloi-Keaikitse, S. (2012). *Classroom assessment practices: A survey of Botswana primary and secondary school teachers* (Unpublished doctoral dissertation). Ball State University, Munchie, Indiana.

Levine, A. (2006). Educating school teachers. Retrieved from the Education Schools Project website http://www.edschools.org/teacher_report_release.htm.

Link, L. (2014). Teachers' perceptions of grading practices (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3648456)

Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Merrill.

Liu, X., O'Connell, A.A., & McCoach, D.B. (2006, April). *The initial validation of teachers' perceptions of grading practices*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), San Francisco, CA.

Lunenburg, F.C. (2010). Schools as open social systems. *Schooling*, *13*(1), 5-11.

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

Marzano, R. J. (2001). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-continent Research for Education and Learning.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results.* Alexandria, VA: Association for Supervision and Curriculum Development.

Maxwell, J.C. (2005). The 360 degree leaders. Developing your influence from anywhere in the organization. Nashville: Thomas Nelson, Inc.

McLoughlin, J.A., & Lewis, R. B. (2008). *Assessing students with special needs* (7th ed.). Columbus, OH: Merrill.

McMillan, J.H. (2007). *Classroom assessment: Principles and practice for effective* (4th ed.). Boston, MA: Pearson Education.

McMillan, J.M., Myran, S., & Workman, D. (2002) Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research, 95*, 203-213.

McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership, 63*(3), 10-17. Retrieved from http://www.ascd.org/ publications/educational-leadership/archived-issues

Middleton, W. (1933). Some general trends in grading procedure. *Education*, *54*(1), 5-10.

Nava, F.J.G., & Loyd, B.H. (1992, April) *An investigation of achievement and nonachievement criteria in elementary and secondary school grading.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Nitko, A.J. (2001). *Educational assessment of students* (3rd. ed.). Upper Saddle River, NJ: Merrill.

O'Connor, K. (2002). *How to grade for learning: Linking grades to standards* (2nd ed.) Thousand Oaks, CA: Corwin.

O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades*. Portland, OR: Educational Testing Service/Assessment Training Institute.

O'Connor, K. (2009). *How to grade for learning: Linking grades to standards* (3rd ed.) Arlington Heights, IL: Skylight.

O'Connor, K. (2012). *How to Grade for Learning by Using 15 Fixes for Broken Grades* Retrieved from http://www.gfps.k12.mt.us/sites/default/files/Ken%20O'Connor%20GFPS%20Aug%2013,%202012.pdf January 22, 2014.

Oliver, J., & Koballa, T. (1992, June). *Science educators' use of the concept of belief.* Paper presented at the meeting of the National Association of Research in Science Teaching, Boston, MA.

Olson, L. (1995, June 14). Cards on the table. *Education Week, 14* (38), 23-28.

Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology, 49*(2), 173-181.

Practitioner. (n.d.). In Merriam-Webster's online dictionary. Retrieved from <http://www.merriam-webster.com/dictionary/practitioner>.

Popham, W. J. (2004). *Classroom assessment: What teachers need to know* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Race to the Top Application for Initial Funding. (2010). Retrieved from http://www.state.tn.us/education/doc/TN_RTTT_Application_2010_01_18.pdf on April 4, 2012.

Reeves, D. B. (2004a). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervision and Curriculum Development.

Reeves, D. B. (2004b). The case against the zero. *Phi Delta Kappan, 86*(4), 324-325.

Reeves, D.B. (2006). Leading to change/ preventing 1,000 failures. *Educational Leadership, 64*(3), 88-89.

Reeves, D.B. (2011). *Elements of grading: A guide to effective practice*. Bloomington, IN: Solution Tree Press.

Rinsland, H. D. (1937). *Constructing tests and grading in elementary and high school subjects*. New York, NY: Prentice-Hall

Rubie-Davies, C.M., Flint, A., & McDonald, L.G. (2011). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology,* 82(2), 270-288.

Russell, C. (1930). *U.S. Patent No. 1772638*. Washington, DC: U.S. Patent and Trademark Office.

Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in Higher Education*, *26*, 327-339.

Silva, M., Munk, D. D., & Bursuck, W. D. (2005). Grading adaptations for students with disabilities. *Intervention in School and Clinic, 41*(2), 87–98.

Starch, K., & Elliot, E.C. (1912). Reliability of grading high school work in English. *School Review, 20: 442-457.*

Stiggins, R. J. (2000). Learning teams for assessment literacy. *Orbit, 30*(4), 5-7.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Portland, OR: Assessment Training Institute.

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

Stiggins, R. J., & Knight, T. (1997). *But are they learning? A commonsense parent's guide to assessment and grading in schools*. Portland, OR: Assessment Training Institute.

Strauss, V. (2014). Everything you need to know about Common Core – Ravitch. *The Washington Post* retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/01/18/everything-you-need-to-know-about-common-core-ravitch/ on August 15, 2015.

Tennessee Department of Education. (2013). Report Card 2012-13. Retrieved from
http://www.tn.gov/education/reportcard/2013.shtml on January 28, 2014.

The Irreplaceables (2012). Understanding the Real Retention Crisis in America's Urban
Schools. Rep. Brooklyn: TNTP, 2012. Retrieved from
http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf on September 13,
2016.

Tomlinson, C. A. (2005, Summer). Grading and differentiation: Paradox or good
practice. *Theory into Practice, 44*(3), 262-269.

TVAAS, (2014).   SAS Institute Inc., Cary, NC, USA. All Rights Reserved. Retrieved
from https://tvaas.sas.com/welcome.html?as=c January, 30, 2014.

Tyack, T., & Tobin, W. (1994). The "grammar" of schooling: Why has it been so hard
to change. *American Educational Research Journal, 31 (3),* 453-479.

U.S. Census Bureau. (1993). *We are Americans: Our education*. Retrieved from http://
www.census.gov/

U.S. Department of Education. (2012). National Center for Education Statistics, Schools
and Staffing Survey (SASS), "Public School Teacher Data File," 2011–12
retrieved from
https://nces.ed.gov/surveys/sass/tables/sass1112_2013314_t1s_002.asp on August
20, 2016.

United States Department of Education. (2010). Race to the Top Assessment Program
Guidance and Frequently Asked Questions. Retrieved from
http://www2.ed.gov/programs/racetothetop-assessment/faqs.pdf on April 6, 2012.

Webster, K.L. (2011). *High school grading practices: Teacher leaders' reflections,
insights, and recommendations* (Doctoral dissertation, Lewis & Clark College).
Retrieved from
http://pqdtopen.proquest.com/pqdtopen/doc/929134496.html?FMT=AI January,
26, 2014.

Wiggins, G. (1993, November). Assessment: Authenticity, context, and validity. *Phi
Delta Kappan, 75*(3), 200-208.

Wiggins, G. (1994). Toward better report cards. *Educational Leadership*, 52 (2), 28-37.

Wiggins, G. (1996). Honesty and fairness: Toward better grading and reporting. In T. R.
Guskey (Ed.), *Communicating student learning: 1996 Yearbook of the association
for supervision and curriculum development* (pp. 141-76). Alexandria, VA:
Association for Supervision and Curriculum Development.

Willis, S. (1993). Are letter grades obsolete? *ASCD Update, 35*(7), 1, 4, 8.

Woofter, T.J. (1917). *Teaching in rural schools*. Boston, MA: Houghton Mifflin.

Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education, 34*(3), 14-27. Retrieved from http://www.stenhouse.com/shop/pc/viewprd.asp?idproduct=8982&r=gb January 26, 2014

Wrinkle, W. L. (1947). *Improving marking and reporting practices in elementary and secondary*. New York, NY: Rinehart.

Zirkle, H. W. (1920, November). Character and results of special rooms as conducted in the Whittier School. *The Elementary School Journal, 21*(3), 189-197.

Zoeckler, L. G. (2007, Spring). Moral aspect of grading: A study of high school English teachers' perceptions. *American Secondary Education, 35*(2), 83-102. Retrieved from http://www.highbeam.com/doc/1P3-1298329441.html January 29, 2014.

Appendix A

Teachers' Perceptions of Grading Practices Survey
Thank you for completing this survey on grading practices. Before you begin, please complete the demographic information in Questions 1-4.
School Information

1. I am (male/female)

2. Age   20-30

             30-40

             40-50

             50-60

3. Years of Experience

         1-5   years

         6-10   years

         11-15   years

         16-20   years

         20-25   years

         25-30   years

         30 and above

4. As part of your college curriculum, have you ever had a class in assessment?

         Yes

         No

Please use the following rating scale as your guide to answering:

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

5. My students' grades are a reflection of my instructional effectiveness.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

6. I consider student effort when I grade.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

7. I will pass a failing student if he or she puts forth effort.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

8. Grades are based on students' completion of homework.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

9. Grades are based on the degree to which students participate in class.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

10. Grades are based on a student's improvement.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

11. Grades are based on students' attendance.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

12. I consider student ability in grading.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

13. Grades are based on students' ability to turn assignments in on time.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

14. Grades are based on students' behavior in class.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

15. Grades are based on students' ability to follow directions.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

16. If a student fails a test, I will offer him/her a second chance to take the test.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

17. If a student fails to complete an assignment, I will assign him/her a grade of zero.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

18. If a student fails to complete an assignment, I will subtract grade points progressively until the assignment is turned in.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

19. I often give students opportunities to earn extra credit.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

20. I often confer with my colleagues on grading criteria.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

21. Grading is the easiest part of my role as a teacher.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

22. It is easy for me to assess student achievement with a single grade or score.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

23. It is difficult to measure student effort.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

24. I need grades to teach well.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

25. Grading can help me improve instruction.

1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree

Appendix B

Hello,
The University of Memphis Institutional Review Board, FWA00006815, has reviewed your submission in accordance with all applicable statuses and regulations as well as ethical principles.

PI NAME: Jeremy Hinson
CO-PI:
PROJECT TITLE: A Correlational Study of Teacher's Perceptions of Grading Practices and their Academic Growth
FACULTY ADVISOR NAME (if applicable): Reginald Green
IRB ID: #3704
CONTINGENCY DATE: 4/6/2016
CONTINGENCY TYPE: Major

The contingencies are listed below:
- Clarify whether this study uses a prospective design or data that has already been collected. Most of the language supports a prospective study, but the investigator continues to provide information about secondary analysis of existing data. It is possible that some data are preexisting while other data are collected prospectively. If the design actually used both prospective data collection and existing data please clarify what is what, how data are linked, and how identity is protected.
- Clarify whether or not the data is not identifiable. Line 50 includes the sentence, "The survey results will remain confidential and if there are any identifiable marks, they will be removed after the data is analyzed." This sentence suggests that there might be identifiable information collected.
- The recruitment email can be written to be more forthcoming with the purpose or question of the study. The current email appears to mask the intention of the study.
- Include both consent document and survey questions with submission.
-Correctly identify your University of Memphis faculty advisor on the IRR form. The advisor should make the affirmation, date and send the revision to the IRB. Once you have addressed the contingencies listed above in your protocol, please e-mail a clean copy of your revised protocol in addition to a copy of your protocol with the changes either highlighted or tracked. These documents need to be e-mailed to irb@memphis.edu with the Subject line Revisions **3704**. If you have a faculty advisor, your faculty advisor will need to be the one who e-mails the revised protocol to our office on your behalf or provide their signature in another fashion in order for the revisions to be processed.

If you have any questions regarding the Board's contingencies, you can contact me via e-mail (irb@memphis.edu). If you have questions regarding how to submit your revised protocol or questions about the IRB process, please contact the Institutional Review Board at irb@memphis.edu or 901-678-2705.
Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair
The University of Memphis