THE CURRICULUM ORIENTATIONS OF SECONDARY STEM TEACHERS AND THEIR BELIEFS ABOUT SOUND GRADING

by

Nathan D. Dean

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in STEM Education

Boise State University

May 2016

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Nathan D. Dean

Thesis Title: The Curriculum Orientations of Secondary STEM Teachers and Their Beliefs About Sound Grading

Date of Final Oral Examination: 01 March 2016

The following individuals read and discussed the thesis submitted by student Nathan D. Dean, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Philip P. Kelly, Ph.D.	Chair, Supervisory Committee
Jennifer L. Snow, Ph.D.	Member, Supervisory Committee
Keith W. Thiede, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Phillip P. Kelly, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

DEDICATION

I dedicate this thesis to my beloved grandpa, Daniel James Doyle. You encouraged me to develop a love of learning and showed me the value of treating others with kindness and respect. Thank you for always believing in me.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the constant love and support of my parents, John and Barb Dean. You never let me lose sight of why I chose this path. Thank you both for pushing me to be my best. I love you.

I would also like to express gratitude for the rest of my family, blood-related and otherwise. Thank you, Dennis, Emily, Joseph, Christa, and Scott. I owe each of you for your encouragement and help, especially when I was feeling overwhelmed.

Finally, I am indebted to my committee members and the rest of the wonderful faculty in the College of Education. Special thanks to Dr. Phil Kelly and Dr. Jennifer Snow; your advocacy, feedback, and uncompromising high standards have helped me grow both professionally and personally. I am eternally grateful for the mentorship and opportunities you have both provided.

ABSTRACT

This study explored two questions: 1) In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices? and 2) How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not? A survey instrument was issued to a sample of secondary science and mathematics teachers (n=89) taken from a metropolitan, northwestern school district. The subsequent analyses showed a significant relationship between the self-actualization orientation and the inclusion of non-academic factors when assigning grades. It was also found that participants who had received formal assessment training were no more likely to endorse literature-recommended grading practices than their untrained colleagues. Implications for assessment training and recommendations for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

STEM	Science, Technology, Engineering, and Mathematics
COI	Curriculum Orientation Inventory
CCSS	Common Core State Standards
NGSS	Next Generation Science Standards

CHAPTER ONE: INTRODUCTION

When examining American education, few issues are as ubiquitous as the assessment and reporting of student achievement. This is especially true given the political push for accountability across our nation's schools and the competitive nature of federal funding programs—recently driven by the No Child Left Behind Act (2002), the Race to the Top Program (U.S. Department of Education, 2009), and the Every Student Succeeds Act (ESSA) of 2015. Though the majority of these accountability measures are based on state-level standardized tests, the heightened emphasis on verifying student achievement has also placed a renewed focus on classroom assessments (DeLuca, 2012). Because of their tangible effects, grades are often the most important product of these assessments from the perspectives of students, parents, and other stakeholders (Airasian & Russell, 2008).

Once grades are assigned, they have a tremendous administrative and emotional effect on the lives of students. Final grades regularly determine ability groupings, class rankings, credits for graduation, serve as the gate keepers for promotion to the next grade-level, and determine admission to post-secondary education. Additionally, grades have a direct influence on student motivation and self-image (Airasian & Russell, 2008; Jung & Guskey, 2011; Randall & Engelhard, 2010; Stiggins, 2007). Given the gravity of what is at stake, it is imperative that teachers report grades that are both valid and reliable.

This raises the question: what distinguishes a teacher as an accurate grader? Previous research by Welsh and D'Agostino (2009) indicates that teachers who adhere to literature-recommended assessment practices are more accurate judges of student achievement than their colleagues who do not (accuracy defined here as the degree to which assigned grades are valid measures of academic achievement). The importance of this finding is underlined by the bulk of studies showing that teachers of all grade levels and content areas regularly violate these recommendations (Brookhart, 1994, 2013a, 2013b; Cross & Frary, 1999; McMillan, 2001; McMillan & Lawson, 2001; McMillan & Nash, 2000; Randall & Engelhard, 2010; Stiggins, Frisbie, & Griswold, 1989). Though a number of external factors contribute to this discrepancy, the most significant internal factor appears to be a teacher's personal philosophy towards teaching and learning (McMillan & Nash, 2000). However, despite these findings, very little is known about the specific traits and beliefs that lead some teachers to adopt recommended grading practices while others do not. This led Brookhart (2013b) to call for additional research focused on determining the precise nature of these qualities.

Research Questions

The purpose of this study was to begin addressing this gap in the literature by determining how various curricular beliefs relate to the acceptance of sound grading practices—specifically with regards to secondary science and math teachers. Using the curriculum orientation framework developed by Eisner and Vallance (1974), I will attempt to answer the following two questions:

1. In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices?

2. How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not?

This document details the research study that I conducted to address these questions. The second chapter is a review of literature relevant to the topic. Following that is a specific description of my methodology. The fourth chapter consists of a detailed analysis of my results. Finally, I conclude with a discussion of my findings and their possible implications.

CHAPTER TWO: LITERATURE REVIEW

The purpose of this review is to provide context and background on the two primary constructs relevant to my study: Curriculum orientation and sound grading. The beliefs that teachers hold with regards to these constructs form the framework of the variables I hope to examine. For the sake of coherence, I have broken down each construct into five component parts—corresponding to the sub-scales of the survey instruments I will use in my study. Details on these instruments can be found in Chapter Three. It should be noted that while secondary science and math teachers are the focus of my study, much of the literature presented here is not content or grade-level specific.

Curriculum Orientation

When it comes to questions of what should be taught and how, the beliefs of teachers are as varied as the individuals who hold them. For this reason, it is useful to adopt a classification system that encompasses the primary trends and assumptions within these beliefs—hereafter referred to as curriculum orientation. Curriculum orientation is defined by Cheung and Wong (2002) as "a collective set of beliefs about curriculum elements such as curriculum intent (aims, goals and objectives), content, teaching strategies, and instructional assessment" (p. 226). They note that while a number of classification schemes have been put forward by scholars, the best-known research on the subject was conducted by Eisner and Vallance (1974), whose framework consisted of five conflicting orientations. Each of these orientations (academic rationalism, cognitive processes, social reconstruction, self-actualization, and technological) are detailed below.

Academic Rationalism

Academic rationalism is the most traditional of the five orientations, centering on the transmission of mankind's greatest works and ideas (Eisner & Vallance, 1974). In this sense, it is heavily focused on developing mental discipline through the study of established subject areas—such as literature, music, mathematics, and science (Jenkins, 2009). Cheung and Wong (2002) note that "rigorous intellectual training" is a major tenant of academic rationalism, requiring students to "act like physicists, historians or mathematicians" (p. 226). Thus, the primary goal is the mastery of academic knowledge and the acquisition of basic skills within the traditional Western disciplines.

Intrinsic to this orientation is the argument that not all knowledge is equally valuable. As such, schools should be selective in the content they transmit. Vocational training and curricula focused on life-skills (e.g., cooking, driving, etc.) tend to be seen as frivolous or even harmful to the true purpose of schooling. To be considered successful through the lens of academic rationalism, the curriculum must prepare students to engage those subjects that most "reflect man's enduring quest for meaning" (Eisner & Vallance, 1974, p. 12).

Cognitive Processes

The cognitive processes orientation holds that the curriculum should focus on developing broadly transferable skills that help students "learn how to learn" (Cheung & Wong, 2002, p. 226). In pursuit of this goal, proponents tend to forgo tasks that require the recall of factual knowledge in favor of those that emphasize problem solving and high-level cognitive skills. This leaves specific subject knowledge as an afterthought, instead highlighting the infinite potential of a student with the skills to be a life-long

learner (Eisner & Vallance, 1974). Jenkins (2009) comments that the "lasting success of this approach is that skills and abilities are not lost when the specific information used to learn the skills or gain the abilities becomes obsolete" (p. 104)

By focusing on the *how* rather than the *what* of learning, a curriculum oriented towards cognitive processes strives to develop independence and adaptability. This stems from the argument that schools cannot control how a student will turn out or what they will do with their education (Eisner & Vallance, 1974). As such, students need to be prepared to apply their thinking skills to the widest range of situations possible. The influence of this orientation can be seen in a number of modern curricular movements, most recently in the development of the Common Core State Standards (CCSS) and Next Generation Science Standards (NGSS). Though neither of these sets of standards do away with traditional content knowledge, both the CCSS and NGSS place a stronger emphasis on developing and assessing cognitive processes than the standards they replace (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; NGSS Lead States, 2013).

Social Reconstruction

A curriculum for social reconstruction attempts to connect academic content to social justice and responsibility. This orientation encompasses a wide spectrum of beliefs, ranging from survivalism (adapting society to meet the needs of a changing world) to radical reformism (preparing students to transform society for the future). In either case, the purpose of a curriculum oriented towards social reconstruction is to develop individuals who can consciously analyze societal issues (Eisner & Vallance, 1974). Advocates of this orientation tend to favor group experiences as well as activities that

require students to engage critically with relevant problems—such as pollution, resource management, and racial discrimination (Cheung & Wong, 2002).

Through the lens of social reconstruction, societal needs take precedence over individual needs. However, it is seen as vitally important that students' interests are addressed and connected to broader social issues (Eisner & Vallance, 1974). In this way, the curriculum can lead young people to develop better habits of mind and become motivated to improve society as a whole (Jenkins, 2009). A curriculum aligned with this orientation can only be considered successful if it inculcates a sense of civic-awareness and critical thought.

Self-Actualization

A curriculum for self-actualization differs strongly by placing the entirety of its focus on the individual student. Described by Eisner and Vallance (1974) as "reconstructionist in a very personal sense" (p. 9), this orientation sees personal growth and fulfillment as the primary goals of the curriculum. In its purest form, this approach calls for education to be a holistic venture that caters to the child's unique interests and emotional needs. Ultimately, proponents of a curriculum for self-actualization want school to be an intrinsically rewarding experience. This approach emphasizes the creation of "an environment where learning is not directed but explored in an open communicative setting which promotes personal growth" (Jenkins, 2009, p. 104).

Like academic rationalism, self-actualization focuses heavily on the content being taught. However, it diverges greatly in the nature of said content. Instead of guiding all students towards a common understanding of the humanity's great works, a curriculum oriented towards self-actualization views the students themselves as a source of direction.

In this way, education can "fulfill its potential as a liberating process by providing integrated experience" (Eisner & Vallance, 1974, p. 9-10). To be deemed effective by proponents of this orientation, the curriculum must not only grow students' cognitive abilities, but also their affective domains (e.g., motivation, self-efficacy, happiness, etc).

Technological

The technological orientation stands apart from the other four in that it claims to be value-neutral. Making no arguments as to *what* should be taught; it focuses exclusively on *how* to teach. This conception views the curriculum as technology—a tool for facilitating learning that can be understood, refined, and implemented. The underlying assumption of this approach is that learning happens in systematic and predictable ways that can be leveraged to increase the effectiveness of instruction for all students (Eisner & Vallance, 1974). Advocates of this orientation focus on finding the most efficient path toward a predetermined set of objectives (whatever those might be). The content is approached in a logical sequence, and instruction tends to be based around recommended practices such as mastery teaching, technology integration, and standards-based assessment (Cheung & Wong, 2002).

Curriculum Meta-Orientation

It should be noted that while these five orientations seem to conflict with one another, there is strong evidence that they are complimentary and interconnected within the minds of teachers (Cheung, 2000; Cheung & Wong, 2002; Jenkins, 2009). This correlation between seemingly disparate belief structures was described by Cheung and Wong (2002) as a curriculum meta-orientation. While some argument has been made that these orientations must be discriminated between and prioritized (Ennis, 1992b), it

appears to be quite possible for teachers to support several orientations simultaneously—even if various pressures force them to enact only one. For this reason, research on teachers' curricular beliefs must avoid treating each orientation as mutually exclusive (Cheung & Wong, 2002). Instead, examining the extent to which teachers hold each orientation independently (as part of their overall meta-orientation) will likely yield better information about how these conceptions interact with their other beliefs.

Currently, there is very little research in that vein. Studies by Cheung (2000), Cheung and Wong (2002), and Jenkins (2009) all focus on the development of quantitative instruments. As such, much of their analyses were dedicated to confirming construct validity. These studies did provide some descriptive findings—such as English teachers supporting a self-actualization orientation more strongly than their Science-teaching colleagues (Cheung & Wong, 2002)—but their scopes were limited. Other research, such as that conducted by Ennis (1992a, 1992b), showed that teachers' curriculum orientations influenced their teaching goals and decisions (barring interfering factors). However, these studies focused exclusively on physical education. To this author's knowledge, there are no existing studies that examine the relationship between teachers' curriculum orientations and their beliefs about assessment or grading.

Sound Grading

Within the professional literature, there is broad agreement as to what constitutes sound grading. Marzano (2010), O'Connor (2009, 2011), Stiggins (2004, 2007), Guskey (2006, 2009, 2011), and others have regularly criticized the failings of traditional grading practices—arguing instead for methods that protect the validity of academic grades as well as the affect of students. Though each author differs slightly in their exact approach

to the problem of grading, a common set of principles unify them. Olson (2013) grouped these principles into five broad categories: The purpose of grades, the inclusion of non-academic factors, the priority of summative vs. formative assessments, methods of calculating and re-calculating grades, and the consistency of grading practices. These categories are explored in greater detail below.

The Purpose of Grades

Brookhart (2011) notes that establishing the purpose of a grade should be the first step in any discussion of grading practices. This is because the validity of any measurement is inextricably tied to its intended purpose (Airasian & Russell, 2008). But what is the purpose of a grade? Amongst measurement specialists, there is general consensus that academic grades should be used solely to communicate a student's level of achievement with regard to the established standards (Allen, 2005; Airasian & Russell, 2008; Bonner & Chen, 2009; Brookhart, 2011, 2013a; Cross & Frary, 1999; Guskey, 2006, 2009, 2011; Marzano, 2010; O'Connor, 2009, 2011; Randall & Engelhard, 2010). It is also generally accepted that the primary audience for these communications are the students themselves and their parents, while secondary consumers of the information include school officials, potential employers, and admissions officers (Allen, 2005; Airasian & Russell, 2008; Brookhart, 2011; O' Connor, 2011; Randall & Engelhard, 2010; Stiggins, 2004).

However, despite the overwhelming agreement found in the literature, teachers and community members vary greatly in their personal beliefs regarding the purposes of assigned grades (Allen, 2005; Brookhart, 2013a; Cross & Frary, 1999; Olson, 2013). The result is that "grades now serve a potpourri of inappropriate purposes including, but not

limited to, self-esteem boosters, public relations, rewards, and vehicles to increase college funding for students" (Randall & Engelhard, 2010, p. 1372). This confusion of purpose leads to what O'Connor (2011) refers to as "broken grades" (p. 136): measurements that are so inaccurate that they become meaningless or even harmful.

Often, this harm is done when grades are used as tools to rank or motivate students. Ranking is done when teachers determine scores by comparing students to one another (rather than pre-established criteria), or by grading "on a curve" (forcing grades to fit a normal distribution). Guskey (2009, 2011) notes that using grades to sort students in this way creates a competitive environment; where one student's success comes at the cost of another student's failure. Furthermore, success in this model does not necessarily mean that a student has achieved anything significant—a student who outperforms his/her peers could still be well below grade-level if they happen to be in a low-achieving cohort group. As for motivation, there is no evidence that the threat of lower grades motivates students to learn more or try harder (Guskey, 2009, 2011). In fact, a punitive approach to grading can serve as a de-motivator for many students (Allen, 2005; O'Connor, 2009, 2011; Stiggins, 2004, 2007). Ultimately, the purpose of grades endorsed by assessment and measurement specialists should be considered more sound than these alternatives.

Inclusion of Non-Academic Factors

The case against the inclusion of non-achievement factors centers around the concept of validity. In essence, a single measurement can only represent one factor at a time while remaining accurate (Guskey, 2006; O'Connor, 2011). This argument is effectively summarized by Allen (2005) when he states that:

> if a teacher must summarize and communicate a student's classroom progress in an academic subject through a single report card grade, then there must be a consensus that the grade represents the most accurate statement of the student's academic achievement, and only academic achievement. This is the essence of valid assessment. To include nonacademic criteria, such as the student's effort, compliance, attitude, or behavior, makes the grade impossible to interpret in any meaningful way. (p. 221-222)

This is not to say that non-achievement factors are unimportant. A study of 488 adolescents by Seider, Gilbert, and Gomez (2012) demonstrated a clear link between academic achievement and character strengths such as perseverance, conduct, and integrity. These findings led them to argue that character factors should be explicitly taught and reported—a sentiment shared by many teachers, who justify their inclusion of non-achievement factors by citing the importance of dispositions and behaviors (Allen, 2005; Randall & Engelhard, 2010).

However, Guskey (2006, 2009) and others have pointed out that assessing and reporting these processes can be done without tainting the accuracy of the academic grade—so long as they are reported separately. Failing to separate these non-achievement factors from a student's academic grade makes it impossible to tell if their reported achievement is the result of actual proficiency or some myriad of other factors like perceived effort and ability (Allen, 2005; Marzano, 2010; O'Connor, 2011). This is especially troublesome since the constructs of effort and ability are difficult to accurately measure, potentially resulting in grades that are biased towards or against certain groups of students (Randall & Engelhard, 2010).

Furthermore, the case for achievement-only grades goes well beyond measurement theory. The way grading is handled directly affects students and their learning (Stiggins, 2007). O'Connor and Wormeli (2011) make the case that diluted

grades cannot support student learning because they are incapable of providing meaningful descriptive feedback. Jung and Guskey (2011) add that the inclusion of non-academic factors can give students the impression that the grades are about who they are, and not about what they can do—a perspective that hinders motivation.

Lastly, there is even some indication that achievement-only grading may have a positive effect on learning outcomes. These results stem from a study of approximately 3800 secondary students in Norway, which controlled for family background, class size, teacher education, and community type. The analysis showed that students exposed to "hard grading" (where high grades were only given for high achievement) outperformed their peers whose grades had been artificially inflated with other factors (Bonesrønning, 2004). Though it's possible that this study may not generalize well to American students, it contributes to a compelling argument for the exclusion of non-academic factors when considered alongside the rest of the literature on grading.

Priority of Formative vs. Summative Assessments

It is also important to consider which measures a teacher *should* use when grading. Unsurprisingly, purpose and validity are once again at the core of this issue. Airasian and Russell (2008) identify two primary uses for assessment: Formative and summative. Formative assessment takes place during the learning process and is used to inform and adjust instruction. This is sometimes referred to as assessment *for* learning (Stiggins, 2004). In contrast, summative assessment is used to determine how successfully a student has learned the material at the end of instruction; also referred to as assessment *of* learning. If the purpose of an academic grade is to communicate a student's ultimate level of achievement, then it follows that scores gathered while instruction in a

topic is still ongoing are less valid for computing final grades than scores gathered after instruction is complete (O'Connor, 2011). Failure to exclude such formative work from a student's final grade can result in an artificially deflated measure of their final achievement—especially if they improve substantially during the learning process.

One of the most common violations of this principle relates the inclusion of homework in final grades. While O'Connor (2011) concedes that homework can be summative (if it entails demonstrating established knowledge), he also points out that homework in America is more often used as practice. Since practice is used to inform learning and develop skills, such homework is essentially formative. For this same reason, Guskey (2006, 2009) considers homework to be a learning "process" and excludes it from the "products" used to determine academic achievement. Sound grading calls for clarity about how an assessment is to be used, and requires that assessments *of* learning be prioritized when calculating academic grades (O'Connor, 2009, 2011).

Methods of Calculating and Re-Calculating Grades

The ways in which grades are calculated (and in some cases, re-calculated) depend on how teachers approach a number of practices like averaging scores, retaking assessments, penalties for late work, and issuing zeroes. As Olson (2013) notes, each of these practices represents a point of contention between traditional approaches to grading and the recommendations for sound grading found in the literature.

Traditional practice often involves averaging all assessments given over a grading period to determine the final grade. However, there are several criticisms leveled against this approach. O'Connor (2011) points out that mean scores are extremely sensitive to outliers, resulting in questionable accuracy if even one assessment is an anomaly. Using a

simple average of all scores during a grading period also fails to acknowledge progress made over time; penalizing students for early failures even if they eventually master the content. Measurement specialists instead recommend considering students' most recent and consistent performances when assigning final grades (Marzano, 2010; O'Connor, 2011; O'Connor & Wormeli, 2011; Stiggins, 2007). This recommendation also validates the practice of allowing students to retake assessments or redo work. If the goal is to measure achievement at the end of instruction, students should be allowed to replace earlier scores if they no longer represent their current capabilities (O'Connor, 2009, 2011).

Late-penalties and zeroes have also been denounced in the literature due to their negative impact on validity. While teachers often attempt to use grade penalties to deter academic dishonesty or poor work ethic, this serves only to confuse the purpose of grades (Allen, 2005). The fact that a student cheats or refuses to do their work does not necessarily communicate anything meaningful about their achievement regarding the course objectives (Guskey, 2009; O'Connor, 2011). In fact, issuing grade penalties for these negative behaviors is simply another way that non-academic factors find their way into academic grades (the problems of which were detailed above). Zeroes are especially problematic since they "give a numerical value to something that has never been assessed" (O'Connor, 2011). Zeroes are also capable of drastically skewing the final grade if scores are averaged (Guskey, 2009; O'Connor & Wormeli, 2011). Instead of late-penalties or zeroes, Guskey (2009) and O'Connor (2011) recommend marking work as "incomplete." Students can then be required to make it up on their own time—encouraging accountability without harming motivation or self-efficacy.

Consistency of Grading Practices

After numerous studies across the country, it goes without question that grading practices differ wildly between teachers. One study by McMillan (2001) examined over 1400 secondary teachers from 53 schools in Virginia. The findings showed that teachers of different content areas lacked agreement on which factors to consider when grading, as well as how those factors should be weighted. Even within content areas, there appears to be little consistency. When McMillan and Lawson (2001) examined the grading practices of 213 secondary science teachers, they found that teachers of low-ability students were less likely to use literature-recommended grading criteria than teachers of higher-ability students. Far from being isolated results, such inconsistencies are common occurrences within the research on grading (Brookhart, 1994, 2013a, 2013b).

These disparities raise significant concerns about reliability—the degree to which assessment information is stable and consistent (Airasian & Russell, 2008). As O'Connor and Wormeli (2011) point out, students achieving at the same level should (ideally) receive the same grade despite having different instructors with different approaches to teaching. For grading practices to be sound, there must be some degree of consistency between individual teachers and schools within a given district (O'Connor, 2011). This is not to say that identical assessments must be used; however, there should be agreement with regards to the expectations and criteria considered when assigning grades.

Discussion

This review examined the literature relating to the two primary constructs investigated by this study: Curriculum orientation and sound grading. Curriculum orientation—as established by Eisner and Vallance (1974)—is a valuable framework for

examining teachers' personal beliefs about teaching and learning. These beliefs are significant because they appear to directly impact practice (Cheung & Wong, 2002; Ennis, 1992a, 1992b). They have also been identified as the primary internal factors influencing how teachers approach grading (McMillan & Nash, 2000). As such, it is reasonable to assume certain curriculum orientations may lead some teachers to be more likely to endorse the principles of sound grading than others. The following chapter details the methodology employed to investigate this relationship in more detail, specifically within the content areas of science and mathematics.

CHAPTER THREE: METHODOLOGY

This research was conducted to explore the relationship between the curriculum orientations of secondary STEM teachers and their beliefs about sound grading. Specifically, the study was organized around the following research questions:

1. In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices?
2. How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not?

The goal of this chapter is to provide a summary of the materials and methods that were used to explore these questions. It begins with a description of the participants, followed by an overview of the survey instruments used. The chapter closes with a description of the overall research design.

Participants

The population of interest for this study consists of secondary STEM teachers within the United States whose primary roles involve teaching science and/or mathematics. Due to limited resources and time restrictions, a convenience sample was drawn from a regional school district. This district—located within a metropolitan area of the northwestern United States—serves over 25,000 students from a variety of socio-economic and cultural backgrounds.

Solicitation

This study sought to obtain anonymous responses from all secondary science and math teachers employed by the selected school district as of Spring 2015. A list of individuals who met these criteria was compiled by reviewing the staff directory page of each school's website. This resulted in a list of 193 teachers from 13 secondary schools who possessed currently-active district email accounts.

After obtaining approval from district and building-level administrators, a solicitation flyer was distributed to the mailboxes of the science and math teachers at each school (Appendix A). The flyer notified potential participants about the upcoming survey distribution as well as the incentive being offered. Participants who completed the survey were given the option to enter a random drawing for one of ten Amazon.com gift cards, ranging in value from $10 to $50.

A solicitation email containing the anonymous survey link was sent to the district accounts of all 193 teachers in my sample (Appendix B). Two reminder emails were also distributed in hopes of improving the response rate. They were sent at one week and two weeks from the initial solicitation, respectively.

Sample and Demographics

Of the 193 teachers solicited, a total of 89 completed the online survey. This equates to a response rate of 46.1%. With regards to gender, 60% of the participants identified as female and 39% identified as male. This compares well to the solicited population, which was 56% female and 44% male. The sample was evenly split by grade-level, with 51% of participants indicating that they taught Junior High School (7^{th} through 9^{th} grade) and 48% teaching Senior High School (10^{th} through 12^{th} grade). In

both of the above cases, a single participant declined to share their demographic information.

Both subject areas were well represented in the sample. Approximately 44% of participants indicated that they primarily taught science, 54% indicated that they taught mathematics, and 2% indicated that their assignment involved teaching both science and math. This is solidly representative of the distribution found in the solicited population (44% science, 56% mathematics). The majority (64%) of the participants had completed a graduate degree of some kind, though this study did not gather more detailed information about the nature of their continuing education. Just over 4% of participants had completed a doctoral degree.

Survey Instrument

This study made use of a three-part survey instrument to gather self-report data from the participant sample (see Appendix D). The first two parts of the survey adapted existing instruments to gather information about the primary constructs in the study: Curriculum orientation and sound grading beliefs. The remaining part consisted of six items designed to collect relevant demographic information.

Curriculum Orientation Inventory (COI)

The first part of the survey utilizes the Curriculum Orientation Inventory (COI) developed by Cheung and Wong (2002). It measures the degree to which participants support each of the curriculum orientations identified by Eisner and Vallance (1974). The instrument is divided into five subscales, corresponding to the five orientations (academic rationalism, cognitive processes, social reconstruction, self-actualization, and technological). Each subscale consists of six items which pose a statement and ask

participants indicate their level of agreement. Participants respond using an 8-point Likert-type scale that ranges from “strongly disagree” to “strongly agree.” Construct validity for this instrument was established using confirmatory factor analysis. The validation study for this instrument reported Cronbach’s α values ranging from 0.77 to 0.83 for each subscale.

Only one significant change was made to the COI for the purposes of this study. A single item from the *academic rationalism* sub-scale was identified by Jenkins (2009) as translating poorly to the United States. The original text of the item reads: “Allowing students to acquire the most important products of our culture is a top priority of the school curriculum.” It was noted that American participants were confused by the term “products” and assumed the item was referencing some form of materialism. For this study, I modified the item to read: “Exposing students to the great works of human culture is a top priority of the school curriculum.” Only minor adjustments were made to the remaining items to ensure that the wording was more natural for an American reader (e.g., “centre” was changed to “center”, etc.).

Sound Grading Questionnaire

The second part of the survey is adapted from an unnamed instrument developed by Olson (2013) as part of a doctoral dissertation at the University of Nebraska. This questionnaire was designed to measure the degree to which educators endorse the sound grading practices recommended by the literature. Like the COI, items in this instrument present statements and ask participants to indicate their level of agreement. It uses a four point Likert-type scale that ranges from “strongly disagree” to “strongly agree.” The survey consists of 22 items and is divided into five subscales: The purpose of grades, the

inclusion of non-academic factors, the priority of summative vs. formative assessments, methods of calculating and re-calculating grades, and the consistency of grading practices.

Olson (2013) worded the majority of items so that agreement corresponds to a rejection of practices put forth by the literature. However, seven items are worded so that greater agreement indicates an endorsement of literature-recommended grading practices. The numerical value of these seven items were reverse-coded before performing the analysis, allowing them to be easily combined with the other items in their subscale.

The face and content validity of this instrument were established by an expert panel. Though no Cronbach's α values were reported by Olson (2013), construct validity and reliability were established using confirmatory factor analysis. No changes were made to this instrument for the purposes of this study.

Data Collection

Responses were collected electronically using online survey software from Qualtrics. Items were presented in three separate pages, corresponding to the three parts of the survey described in the previous section. Each part is identical to what is found in Appendix D, with the exception that the items in Part 1 (Curriculum Orientation) and Part 2 (Sound Grading Beliefs) were randomized to obscure the relationship between items within the same subscale.

Participant Protections

The solicitation emails contained a generic link (generated by Qualtrics) to ensure that individually identifiable data was not collected. To guarantee that participants were informed of their rights and the nature of the study before participating, the solicitation

emails contained a brief description of the research goals and linked directly to the informed consent page (see Appendix C). Participants were not given access to the survey until they indicated that they had provided consent. Following data collection, individual responses were stored within a password-secured, university-issued cloud account.

It was also necessary to ensure that participation in the incentive offer did not interfere with the anonymity of the data collection. To aid in this, entry into the drawing was handled outside of Qualtrics. The final page of the survey informed participants that they could enter the drawing by sending a message from the email account of their choice to my university-issued email address. It was not necessary to include additional identifiable information in the email since the Amazon.com gift cards were distributed directly back to the email addresses that were used to enter the drawing. Because there are no direct links between the Qualtrics survey and the participants' emails regarding the incentive, the anonymity of individual responses was satisfactorily maintained.

Methods of Analysis

Following data collection, the responses were analyzed. A quantitative design was used to explore the research questions of the study. An overview of the statistical methods used to answer each question is provided below.

Research Question 1

The first question that this study sought to answer was, "In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices?" Implicit here is the hypothesis that a correlation between the variables in question exists. To determine if this is the case, a correlational analysis was performed

to discover whether or not a significant relationship could be found between the curriculum orientation subscales and each of the sound grading subscales.

Research Question 2

This study also attempted to address the question, “How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not?” To answer this question, a series of independent-samples t-tests were performed to compare the responses of participants who had participated in formal assessment training with those who indicated that they had not. These analyses compared responses across each of the sound grading subscales. The associated data and results are presented in the next chapter.

CHAPTER FOUR: RESULTS

The primary goal of this study was to explore the curriculum orientations of secondary STEM teachers and their beliefs about sound grading. The purpose of this chapter is to present the data gathered, as well as the results of my analyses. Prior to carrying out any statistical tests, a reliability analysis was performed to verify that the internal consistency of each subscale was sufficient for further use. Subsequently, a series of statistical tests were performed to uncover significant relationships or differences present within the data. SPSS Version 21.0 was used to perform all tests, the results of which are organized according to the research questions that they address.

Reliability Analysis

The two primary constructs examined by this study are *curriculum orientations* and *sound grading beliefs*. The survey instruments used to measure these constructs divided each one into five subscales that represent important themes found in the existing literature (see Appendix D). Cronbach's α was calculated for each subscale to verify that the items within those subscales had sufficient internal consistency to justify their use in further statistical tests.

Initially, I calculated the α values using responses from all 89 participants. However, several participants had left individual items blank, resulting in missing data. To determine if it would be more beneficial to consider only responses from those with complete data, Cronbach's α was recalculated after removing participants who had failed to provide a response on all items (n=71). The values found in both instances are

provided in Table 1. As there were no substantial differences in the reliability values of any subscale, I decided to continue my analysis using the responses from all participants (even those with missing data) to maximize the effective sample size and statistical power.

Table 1: Values for Cronbach's α by Subscale

	All Participants Included (n = 89)	*Participants with Missing Data Removed (n = 71)*
Curriculum Orientation Subscales		
Academic Rationalism	**0.52**	0.53
Cognitive Processes	**0.59**	0.59
Social Reconstruction	**0.87**	0.84
Self-Actualization	**0.65**	0.61
Technological	**0.78**	0.80
Sound Grading Subscales		
Purpose of Grades*	**0.32**	0.28
Non-Academic Factors	**0.56**	0.59
Formative vs. Summative*	**0.33**	0.35
Methods of Calculation	**0.74**	0.77
Consistency	**0.80**	0.80

*Subscales removed from further analyses due to extremely low internal-consistency

Many of the subscales possessed values for Cronbach's α that were lower than expected, indicating that they could lack the internal consistency necessary for valid interpretation. The *Purpose of Grades* and *Formative vs. Summative* subscales possessed the lowest α values, both well below 0.5. Since Cronbach's α is sensitive to the number of items present (Gliem & Gliem, 2003), it is possible that these exceptionally low values stem from the relatively few items in these subscales (three and two respectively). The remaining subscales each contained between four and seven items. However, given the

difficulty in defending their use, I decided it would be best to remove the *Purpose of Grades* and *Formative vs. Summative* subscales from further analyses.

Several of the remaining subscales also possessed values for Cronbach's α that are problematic. However, I decided to continue including them in my analyses for two reasons. First, while α values of 0.5 to 0.6 are considered poor, they avoid being classified as unacceptable by the widely referenced rules of thumb put forth by George and Mallory (as noted in Gliem & Gliem, 2003). Second (as can be seen in Table 2), previous studies utilizing the Curriculum Orientation Inventory (COI) have reported substantially higher α values for the subscales in question (Cheung & Wong, 2002; Jenkins, 2009). Since those studies made use of much larger sample sizes, it stands to reason that their measurements of internal consistency may be more telling than the ones obtained in this study.

Table 2: Reported Cronbach's α in Studies Utilizing the COI

	Cheung & Wong, 2002 (n = 648)	*Jenkins, 2009 (n = 308)*	***Current Study (n = 89)***
Academic Rationalism	0.78	0.66	**0.52**
Cognitive Processes	0.77	0.61	**0.59**
Social Reconstruction	0.83	0.74	**0.87**
Self-Actualization	0.78	0.85	**0.65**
Technological	0.79	0.76	**0.78**

Despite these justifications, the issue of internal consistency cannot be ignored completely. In the seminal text *Psychometric Theory*, Nunnally and Bernstein (1994) suggest that α values below 0.70 are undesirable—even in the early stages of research. Though I have chosen to carry out my analyses in spite of this, the problematic internal

consistency of my remaining subscales highlights serious and unavoidable limitations to the interpretation of my results.

How Are Curriculum Orientations Predictive of Sound Grading Beliefs?

The first question that this study sought to answer was "In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices?" To explore this question, I performed a series of partial correlations to examine the relationships between each of the five curriculum orientation subscales and the three remaining sound grading subscales (see Table 3). The correlations were controlled for the effects of gender, content-area, grade-level, education level, and years of teaching experience. Categorical responses (such as gender) were dummy-coded to allow them to be factored into the analysis.

Table 3: Correlations Between Curriculum Orientations and Grading Beliefs

n=72 Correlation (P-Value)	***Non-Academic Factors***	***Methods of Calculation***	***Consistency***
Academic Rationalism	0.13 (0.26)	0.12 (0.31)	-0.05 (0.65)
Cognitive Processes	0.02 (0.88)	-0.02 (0.14)	-0.14 (0.23)
Social Reconstruction	0.20 (0.09)	-0.09 (0.46)	-0.07 (0.57)
Self-Actualization	**0.27 (0.02)***	-0.10 (0.38)	-0.08 (0.52)
Technological	-0.04 (0.71)	-0.02 (0.85)	0.01 (0.91)

* Significant at the 0.05 level

The only significant relationship found was between the *self-actualization* subscale and the *non-academic factors* subscale (correlation=0.27, p=0.02). The positive correlation indicates that participants who espoused a self-actualization orientation were also more likely to endorse the inclusion of non-academic factors when assigning grades (in contrast to the recommendations found in the literature).

How Do Beliefs Differ with Regards to Formal Assessment Training?

The second question that drove this study was, "How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not?" To address this question, I performed a series of independent-samples t-tests to compare the subscale scores of those who reported that they had received assessment training with those who reported that they had not.

Although the survey instrument used in this study gathered ratio data regarding the hours of formal training that participants had received, it should be noted that I decided to treat the data categorically for the purposes of my analyses. This is due to several responses containing values that stood out as potentially invalid (e.g., indicating over 100 hours of formal assessment training, despite being a first year teacher without an advanced degree). While it is theoretically possible that these responses accurately represent the experiences of the participants, it seems much more likely that the scores had been inflated—either by self-report bias or a misunderstanding of what qualified as formal training for the purposes of this study. Using a binary comparison allowed me to eliminate any influence from potentially inflated responses while still presenting the full, unaltered data set. See Table 4 for the results of my comparison between those who had received formal assessment training and those who had not.

Table 4: Differences in Grading Beliefs with Regards to Assessment Training

	No Training (n=6) Mean / Std. Dev.	***Some Training (n=78)*** Mean / Std. Dev.	***P-Value***
Non-Academic Factors	1.97 / 0.56	2.16 / 0.52	0.40
Methods of Calculation	2.43 / 0.69	2.53 / 0.58	0.70
Consistency	**2.04 / 0.80**	**2.73 / 0.72**	**0.03***

* Significant at the 0.05 level

Participants who had received assessment training only differed significantly with regards to the *consistency* subscale of the sound grading questionnaire ($p=0.03$). In this case, their higher mean score indicates that those with training were less likely to support the establishment of uniform grading criteria than those without training. This shows that participants who had not received formal assessment training were more in-line with the literature-based recommendations than those who had been trained.

Summary

This study sought to examine the curriculum orientations of secondary STEM teachers, as well as their beliefs about sound grading practices. When examining the data, both of the questions guiding this research had statistically significant findings. The results and conclusions drawn from these analyses are discussed in the following chapter.

CHAPTER FIVE: DISCUSSION

In this study, I sought to examine the curriculum orientations and grading beliefs of secondary STEM teachers. Specifically, I gathered data to address the following research questions:

1. In what ways are the curriculum orientations of STEM teachers predictive of their beliefs regarding sound grading practices?
2. How do STEM teachers who have received formal training in assessment differ in their grading beliefs from those who have not?

I explored these questions by collecting responses from secondary STEM teachers in a metropolitan, northwestern school district using a 58-question survey instrument. My subsequent quantitative analyses found significant results for both questions. The purpose of this chapter is to provide a discussion of my results (arranged by research question) and their implications. The chapter closes with an outline of the limitations of this study as well as a list of recommendations for future research.

How Are Curriculum Orientations Predictive of Sound Grading Beliefs?

This study found that one of the orientations proposed by Eisner and Vallance (1974) was predictive of secondary STEM teachers' beliefs about sound grading (correlation=0.27, p=0.02). Specifically, participants who endorsed the self-actualization orientation were also more likely to support the inclusion of non-academic factors when assigning grades. This result resonates with previous studies, which found that teachers often attempt to include non-academic factors in the gradebook for the purpose of

motivating or enabling students (Bonner & Chen, 2009; Brookhart, 2013a; Cross & Frary, 1999; McMillan, 2001; McMillan & Nash, 2000; Randall & Engelhard, 2010).

As detailed in Chapter Two, the self-actualization orientation focuses on the intrinsic motivation, emotional well-being, and holistic growth of students as individuals. Thus, it seems natural that a strong endorsement of this orientation would result in an equally holistic approach to grading. Given the findings of previous studies (Bonner & Chen, 2009; Brookhart, 2013a; Cross & Frary, 1999), this is almost certainly done with good intentions—such as the desire to provide encouragement or foster personal growth in areas that are not strictly academic. However, the likely outcome is that academic grades are diluted to the point that they lack any meaningful validity.

The problems associated with this approach to grading are not limited to issues of measurement theory. Ironically, assigning grades that consider non-academic factors may undermine the primary goals associated with the self-actualization orientation. As O'Connor (2011) states, "Intrinsic motivation is clearly in conflict with the use of grades as extrinsic motivators" (p. 8). This idea is expanded upon by Jung and Guskey (2011), who describe the negative effect this conflict can have:

> When students' grades are inflated and not clearly connected to achievement on well-defined outcomes, students begin to believe that grades are not about what they do, but about who they are. Such adaptations to grades actually lead to a decrease in motivation. (p. 34)

These statements highlight a contradiction between the underlying motivations of STEM teachers' grading beliefs and the actual outcomes of their implementation. If educators hope to produce valid grades *and* support the affective growth of students, a clear separation must be made between academic achievement and non-academic traits.

Also of interest are the relationships not found in my analyses. Though the connection between the self-actualization orientation and the inclusion non-academic factors is potentially revealing, it is surprising that no other significant correlations were found in my data. While it is impossible to conclusively determine the reason for this, I see two plausible explanations: 1) Additional relationships exist but the subscales of my instrument did not hold together well enough to reveal them, or 2) The effects of participants' orientations were overpowered by other factors.

The first scenario seems quite likely, given the problematic internal consistency of my instrument. If the items within a given subscale are not well correlated, it only makes sense that relationships involving that subscale would be difficult to discern. Since previous research seems to support the idea that curriculum orientations should strongly correlate with grading beliefs, it is probable that better instrumentation would likely reveal relationships that were not identified in this study. Given its strong standards-based approach and adherence to research-based methods, I would especially expect to see a correlation between the technological orientation and the endorsement of literature recommended grading practices.

The second scenario is also reasonable. It is well known that external factors can have a strong influence on a teacher's grading beliefs and practices. One such confounding factor may be the influence of tradition. As Jung and Guskey (2011) state, "most current grading practices are grounded in tradition, rather than research on best practice" (p. 32). A teacher's grading beliefs will only correlate with their curriculum orientation if those beliefs have been carefully aligned with their overall educational philosophy. In the absence of such intentional strides toward philosophical coherence,

tradition may wield greater influence by default. Therefore, the apparent disconnect between curriculum orientation and sound grading beliefs found in my study may be partially explained by participants endorsing what they know and are comfortable with—without consideration of whether it aligns with their dominant orientations.

How Do Beliefs Differ with Regards to Formal Assessment Training?

A study by Cross and Frary (1999) found that teachers who had received substantial training in assessment were just as likely to violate recommended grading practices as those with no training at all. My data strongly support their conclusions.

The results of this study show that secondary STEM teachers who have received assessment training are no more likely to endorse literature-recommended grading practices than those without training. Not only were trained teachers equally supportive of including non-academic factors or calculating grades in ways that contradicted recommendations, they were—surprisingly—even more opposed to the establishment of consistent grading practices than their untrained colleagues ($p=0.03$).

Although this study did not examine the details of the formal training that participants had received, these results present some interesting implications. At face value, the data seems to imply that the reported assessment training was ineffective (or even harmful). However, without knowing the professional development context of the district that my sample was drawn from, it is difficult to rule out potential confounding factors. For instance, it is possible that professional development focused on other topics may have reinforced some of the problematic grading beliefs that this study sought to measure. In such a case, the lack of difference found between those with assessment

training and those without could be attributed to conflicting messages rather than problems inherent to the assessment training itself.

However, when considered alongside previous studies that showed little effect from measurement training (Brookhart, 1994; Cross & Frary, 1999), these results begin to call into question the efficacy of current approaches to preparing teachers in assessment. This is especially true when viewed in light of the abundance of other studies showing that teachers continue to assign grades in ways that undermine their validity. While the need for more training in educational measurement is clearly established in the literature (Allen, 2005; Brookhart, 1994; Jung & Guskey, 2011), more may not be enough on its own. To elicit meaningful changes in beliefs about grading, assessment training may need to refocus on addressing the core philosophies that lead teachers to endorse problematic grading practices to begin with.

Conclusion

The link between the self-actualization orientation and problematic grading beliefs provides some insight as to how training in assessment might be improved. If the philosophy underlying the inclusion of non-academic factors centers on non-technical elements (like a student's affective growth and well-being), it is doubtful that validity-based arguments will carry much weight on their own. Teachers who include non-academic factors are unlikely to change their grading beliefs unless they perceive a benefit that aligns with their dominant curriculum orientations. Though this study focused on secondary STEM teachers, it seems unlikely that this conclusion would be exclusive to specific content areas.

As such, I contend that assessment training for all teachers may be improved by addressing the self-actualization aspects of their meta-orientations directly. This does not imply that instruction on the technical aspects of educational measurement should be minimized, however. Instead, my assertion is that training in assessment must also demonstrate how individual students are better served by the use of literature-recommended grading practices. Ultimately, teachers align their beliefs and practices with what they think is best for the children they teach. The conflict between including non-academic factors in grades and supporting the affective growth of students needs to be made explicit in order for training programs to create a meaningful change in grading beliefs. However, development of a specific plan for implementing this concept is beyond the scope of this study.

Limitations

One of the limitations of this study is its reliance on self-report data. As Gay, Mills, and Airasian (2006) note about self-report instruments, "the researcher can never be sure that individuals are expressing their true attitude, interest, values, or personality, as opposed to a 'socially acceptable' response" (p. 132). Additionally, responses to a self-report instrument may be biased towards what the participants believe the researcher hopes to find (Gay et al., 2006). It is possible that the responses collected are skewed in favor of what the participants thought would be desired or acceptable. This is especially true for the amounts of formal assessment training that were reported. However, the binary treatment of this data helps limit the effects of potential bias, and the results of the corresponding analysis were well-aligned with previous findings.

Another potential limitation stems from the sampling method employed. Because this study relied on a convenience sample, the results may not be perfectly representative of the broader population of American secondary STEM teachers. Gay et al. (2006) state that:

> Because the total population is composed of both volunteers and nonvolunteers, the results of a study based solely on volunteers are not likely to be generalizable to the entire population. (p. 112)

My convenience sample is also limited by being drawn from a single school district in the northwest, which may cause responses to differ greatly from what would be found in other contexts. For instance, the unique professional development context of this district will have undoubtedly influenced my results.

Finally, though the subscales with the lowest internal consistency were removed, some of the remaining subscales had Cronbach's α values that are widely considered to be poor. This has undoubtedly harms the strength of conclusions drawn from the associated results. However, given the alignment of my findings with those of previous studies, I am reasonably confident that my conclusions can provide a modest basis for further exploration.

Recommendations for Future Research

Based upon the results and limitations of this study, there are two areas which I contend would benefit from further research. First, effort should be made to refine or develop new survey instruments for investigating both curriculum orientations and beliefs about sound grading. The limited availability of such instruments—and the inconsistent reliability of their subscales—shows that this is an area ripe for development.

Second, although the conclusions drawn in this study are potentially impactful, they are based on a very constrained data set with serious methodological limitations. Future research could focus on validating my findings with more robust sampling and instrumentation. Alternatively, assessment training that explicitly addresses the self-actualization orientation could be developed and piloted to test my conclusions empirically.

REFERENCES

Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment: Concepts and Applications*. Boston: McGraw-Hill.

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

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

Bonner, S. M., & Chen, P. P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, *14*(2), 57-77.

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

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

Brookhart, S. M. (2013a) Grading. In J. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 257-274). Sage Publications.

Brookhart, S. M. (2013b). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 69-90.

Cheung, D. (2000). Measuring teachers' meta-orientations to curriculum: Application of hierarchical confirmatory factor analysis. *The Journal of experimental education*, *68*(2), 149-165.

Cheung, D., & Wong, H. W. (2002). Measuring teacher beliefs about alternative curriculum designs. *Curriculum journal*, *13*(2), 225-248.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied measurement in Education, 12*(1), 53-72.

DeLuca, C. (2012). Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education. *Action in Teacher Education*, *34*(5-6), 576-591.

Eisner, E. W., & Vallance, E. M. (Eds.). (1974). *Conflicting conceptions of curriculum*. McCutchan Publishing.

Ennis, C. D. (1992a). Curriculum theory as practiced: Case studies of operationalized value orientations. *Journal of Teaching in Physical Education, 11*(4), 358-375.

Ennis, C. D. (1992b). The influence of value orientations in curriculum decision making. *Quest*, *44*(3), 317-329.

Every Student Succeeds Act of 2015, S. 1177, 114th Cong. (2015)

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Education research: Competencies for analysis and applications* (8th ed.). Columbus, OH: Pearson Education.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, Ohio*, 82-88.

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

Guskey, T. R. (Ed.).(2009). *Practical solutions for serious problems in standards-based grading.* Thousand Oaks, CA: Corwin Press.

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

Jenkins, S. B. (2009). Measuring teacher beliefs about curriculum orientations using the modified-curriculum orientations inventory. *The Curriculum Journal*,*20*(2), 103-120.

Jung, L. A., & Guskey, T. R. (2011). Fair & accurate grading for exceptional learners. *Principal Leadership, 12*(3), 32-37.

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

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20-32.

McMillan, J. H., & Lawson, S. (2001). *Secondary science teachers' classroom assessment and grading practices.* ERIC Document 450 158.

McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Authors.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states.* Washington, DC: The National Academies Press.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

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

O'Connor, K. (2011). *A repair kit for grading: 15 fixes for broken grades*. Pearson.

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

Olson, M. E. (2013). *Making the Grade: Do Nebraska Teachers and Administrators Working in Public Schools in 7th--12th Grade Settings Agree About What Constitutes Sound Grading Practice?* (Doctoral dissertation). University of Nebraska.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380.

Seider, S., Gilbert, J. K., & Gomez, J. (2012). The role of moral and performance character strengths in predicting achievement and conduct among urban middle school students. *Teachers College Record*, *115*, 1-18.

Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, *86*(1), 22-27.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership, 64*(8), 22.

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

U.S. Department of Education. (2009). *Race to the Top Program Executive Summary*. Washington, D.C.. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Welsh, M. E. & D'Agostino, J. V. (2009). Fostering consistency between standards-based grades and large-scale assessment results. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 75-104). Thousand Oaks, CA: Corwin Press.

APPENDIX A

Solicitation Flyer

MATH AND SCIENCE TEACHERS WANTED FOR A RESEARCH STUDY

The Relationship between the Curriculum Orientations of Secondary STEM Teachers and their Beliefs about Sound Grading

Are you a secondary math or science teacher?

We are conducting a research study to determine if the curricular beliefs of science and math teachers are predictive of their beliefs about grading ... And we would like your input!

--- --- ---

Description: Volunteers will be asked to take an online survey. Completing the entire survey will require approximately 15 minutes.

Eligibility: To participate, you must be a certified teacher who currently teaches math and/or science at the secondary level.

Incentive: As a thank you, participants will be eligible for a drawing to receive one of ten Amazon.com gift certificates – ranging in value from $10 to $50.

How to Participate: Just keep an eye on your district email account!
An email containing a link to the survey will be distributed in the next day or two.

--- --- ---

This research is being conducted by Nathan Dean under the advisement of Dr. Philip Kelly (Department of Curriculum, Instruction, and Foundational Studies – Boise State University)

APPENDIX B

Solicitation Email

Greetings.

My name is Nathan Dean and I am a graduate student in the Department of Curriculum, Instruction, and Foundational Studies (CIFS) at Boise State University. As part of my master's thesis, I am conducting a research study to determine if the curricular beliefs of science and math teachers are predictive of their beliefs about grading.

I am writing to ask if you would be willing to take about 15 minutes to complete a survey for this research project. Participation is completely voluntary and your answers will be anonymous.

As a thank you, participants will be eligible for a drawing to receive one of ten Amazon.com gift certificates – valued at $10 to $50.

If you are interested, please click the link below for the survey and additional information:

[Survey link here]

If you have any questions, please do not hesitate to contact me (natedean@u.boisestate.edu) or my advisor, Dr. Philip Kelly (pkelly@boisestate.edu).

I greatly appreciate your time. Every response is incredibly valuable for the success of this study.

Thank you,

Nathan Dean
Graduate Student
College of Education
Boise State University

APPENDIX C

Informed Consent Page

BOISE STATE UNIVERSITY

The Relationship between the Curriculum Orientations of Secondary STEM Teachers and their Beliefs about Sound Grading

Nathan Dean, a graduate student at Boise State University, is conducting a research study to determine if the curricular beliefs of science and math teachers are predictive of their beliefs about grading. You are being asked to complete this survey because you currently teach one of these subjects at the secondary level.

Participation is voluntary. The survey should take approximately 15 minutes to complete.

This study involves no foreseeable serious risks. We ask that you try to answer all questions; however, if there are any items that make you uncomfortable or that you would prefer to skip, please leave the answer blank. Your responses are anonymous.

If you have any questions or concerns feel free to contact Nathan or his faculty advisor:

Nathan Dean, Graduate Student
College of Education
Boise State University
(208) 371-1756
natedean@u.boisestate.edu

Dr. Philip Kelly, Professor
College of Education
Boise State University
(208) 426-4977
pkelly@boisestate.edu

If you have questions about your rights as a research participant, you may contact the Boise State University Institutional Review Board (IRB), which is concerned with the protection of volunteers in research projects. You may reach the board office between 8:00 AM and 5:00 PM, Monday through Friday, by calling (208) 426-5401 or by writing: Institutional Review Board, Office of Research Compliance, Boise State University, 1910 University Dr., Boise, ID 83725-1138.

If you would prefer not to participate, please do not fill out a survey.

If you consent to participate, please complete the survey.

APPENDIX D

Survey Instrument

Part 1: Curriculum Orientation

INSTRUCTIONS: Please read each of the following statements carefully, and indicate your level of agreement on a scale of 1 (Strongly Disagree) to 8 (Strongly Agree).

Answer according to what you personally believe is best. This may or may not reflect the actual approach to curriculum in your school or district.

Please try to answer every item. If unsure, respond to the best of your ability.

Note: Item order was randomized to obscure subscales (this note was not included in the survey distributed to participants)

Academic Rationalism

- Exposing students to the great works of human culture should be a top priority of the school curriculum.
- The most important curriculum content for primary and secondary school students should be subject knowledge.
- Curriculum should require teachers to transmit the best and most important subject matter to students.
- It is important to assess the extent to which students have acquired basic subject knowledge.
- Subject knowledge is the basis for designing high-quality school curriculum.
- Curriculum should stress refinement of students' intellectual abilities.

Cognitive Processes

- The basic goal of curriculum should be the development of students' cognitive skills, such as memorizing, hypothesizing, problem-solving, analyzing and synthesizing, which can be applied to learning virtually anything.
- Methods of inquiry are the most important content for primary and secondary school curricula.
- Curriculum should first let students master cognitive skills (e.g. deducing, analyzing, critical thinking) and then teachers may teach conceptual knowledge.
- During the teaching–learning process, it is most important to give students opportunities to think about problems.

- Assessing students' levels and forms of thinking as well as their ability to explore knowledge is most important.

- Curriculum should require teachers to teach thinking skills systematically.

Social Reconstruction

- Assessment of students should emphasize civic awareness, problem-solving skills, and decision-making skills.

- Students learn best when permitted to analyze, investigate, and evaluate authentic societal problems.

- Existing problems in our society, such as pollution and population explosion, should be the organizing center of the curriculum.

- Curriculum content should focus on societal problems such as pollution, population explosion, energy shortage, racial discrimination, and crime.

- The most important goal of the school curriculum is to foster students' abilities to critically analyze societal problems.

- The curriculum should let students understand societal problems and take action to establish an improved society.

Self-Actualization

- Teachers should select curriculum content based on students' interests and needs.

- Students' interests and needs should be the organizing center of curriculum.

- During the teaching process, teachers should frequently check whether students are provided with opportunities to integrate their affective, cognitive, and psychomotor developments.

- Students learn best in a learning environment filled with love and emotional support.

- In addition to academic achievements, instructional assessment should also emphasize students' personal development such as self-confidence, motivation, interests, and self-concept.

- Curriculum should try to provide satisfactory learning experiences for each student.

Technological

- Selection of curriculum content and teaching activities for every school subject should be based on the established learning objectives.
- Curriculum organization should be governed by the ordering of the learning objectives.
- Teaching should focus on finding efficient means to reach a set of predetermined learning objectives.
- For curriculum design, the main function of instructional assessment is to find out the extent to which students have attained the intended learning objectives.
- Learning occurs in certain systematic ways.
- Curriculum design should start with stating learning objectives.

Part 2: Sound Grading Beliefs

INSTRUCTIONS: Please read each of the following statements carefully, and indicate your level of agreement on a scale of 1 (Strongly Disagree) to 4 (Strongly Agree).

Answer according to what you believe constitutes best practice in grading. For a variety of reasons, this may or may not reflect your actual practice.

Please try to answer every item. If unsure, respond to the best of your ability. To avoid confusion, certain words have been underlined in some statements.

Note: Item order was randomized to obscure subscales, (*) indicates reverse coded items (this note was not included in the survey distributed to participants)

The Purpose of Grades

- *The primary purpose of grades is to communicate what a student has learned.
- The purpose of a grade is to show how students are achieving relative to one another.
- There should be a limit to the number of students whose final grade is an “A.”

Inclusion of Non-Academic Factors

- Attendance should be a component of a final grade.
- Tardiness should be a component of a final grade.
- A student's effort should be taken into account when assigning a final grade.
- Grades should be used to reward students who work hard.
- A student's ability should be taken into account when assigning a final grade.
- Extra-credit work is an appropriate way to help low achieving students reach a passing grade.

Priority of Formative vs. Summative Assessments

- Performance on homework should be a component of a final grade.
- *Teachers should use grades from summative measures rather than formative measures when determining a student's final grade.

Methods of Calculating and Re-Calculating Grades

- Points should be deducted from work submitted late.
- Zeroes should be given for cheating.
- Final grades should include zeroes for incomplete or missing work.
- Students should not be permitted to redo work or retake tests.
- *Students must redo work or retake tests if they have not mastered the content.
- *A higher score on a retake should fully replace the initial lower score.
- Averaging every score given over the course of a term is the best way to determine a final grade.

Consistency of Grading Practices

- Each teacher should develop and use his or her own individual grading practices.
- *All teachers in the same building and in the same discipline/subject area should use the same grading practices.

- *All teachers in the same building, regardless of discipline/subject area, should use the same grading practices.

- *All teachers across a district, regardless of discipline/subject area, should use the same grading practices.

Part 3: Demographics

INSTRUCTIONS: Please answer the following items to the best of your ability.

- Please indicate your gender
  (Female, Male)

- How many years have you been teaching?
  (Numerical Response)

- What is the highest level of education that you have obtained?
  (Bachelor's Degree, Master's Degree, Doctoral Degree)

- In your current assignment, what grade-level do you primarily teach?
  (Junior High School, Senior High School)

- In your current assignment, what content area do you primarily teach?
  (Science, Math, Both Science and Math, Neither Science or Math)

- Approximately how many hours of formal training have you received in assessment or educational measurement? *Formal training can include school or district professional development, conference training, professional learning communities (PLCs), or college coursework. If you have taken college coursework in assessment, convert the course credits to hours using this ratio: 1 credit = 15 hours. (Example: A 3-credit college course in assessment would be worth 45 hours for this response).*
  (Numerical Response)

APPENDIX E

IRB Notification of Exemption

BOISE STATE UNIVERSITY
RESEARCH AND ECONOMIC DEVELOPMENT

Date: April 16, 2015

To: Nathan Dean cc: Phillip Kelly

From: Office of Research Compliance (ORC)

Subject: SB-IRB Notification of Exemption - 108-SB15-079

The relationship between the Curriculum Orientations of Secondary STEM Teachers and their Beliefs about Sound Grading

The Boise State University ORC has reviewed your protocol application and has determined that your research is exempt from further IRB review and supervision under 45 CFR 46.101(b).

Protocol Number: 108-SB15-079
Approved: 4/16/2015 Application Received: 4/14/2015
Review: Exempt
Category: 2

This exemption covers any research and data collected under your protocol as of the date of approval indicated above, unless terminated in writing by you, the Principal Investigator, or the Boise State University IRB. All amendments or changes (including personnel changes) to your approved protocol **must** be brought to the attention of the Office of Research Compliance for review and approval before they occur, as these modifications may change your exempt status. Complete and submit a Modification Form indicating any changes to your project.

Annual renewals are not required for exempt protocols. When the research project is completed, please notify our office by submitting a Final Report. The exempt status expires when the research project is completed (closed) or when the review category changes as described above.

All forms are available on the ORC website at http://goo.gl/D2FYTV

Please direct any questions or concerns to ORC at 426-5401 or humansubjects@boisestate.edu.

Thank you and good luck with your research.

Office of Research Compliance

1910 University Drive Boise, Idaho 83725-1139
Phone (208) 426-5401 orc@boisestate.edu