8-2019

Standards-Based Grading Practices in Middle School Mathematics Classrooms: A Multicase Study

Michelle Ann Morgan
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

STANDARDS-BASED GRADING PRACTICES IN MIDDLE SCHOOL MATHEMATICS CLASSROOMS: A MULTICASE STUDY

A Dissertation Submitted in Partial Fulfillment
Of the Requirements for the Degree of
Doctor of Philosophy

Michelle Ann Morgan

College of Natural and Health Sciences
School of Mathematics Sciences
Educational Mathematics

August 2019

This Dissertation by: Michelle Ann Morgan

Entitled: *Standards-Based Grading Practices in Middle School Mathematics Classrooms: A Multicase Study*

has been approved as meeting the requirements for the Degree of Doctor of Philosophy in the College of Natural and Health Sciences in the School of Mathematical Sciences, Program of Educational Mathematics.

Accepted by the Doctoral Committee

____

Dr. Robert A. Powers. Ed.D., Research Advisor

____

Dr. William L. Blubaugh, Ph.D., Committee Member

____

Dr. Steven C. Leth, Ph.D., Committee Member

____

Dr. Mark A. Smith, Ph.D., Faculty Representative

Date of Dissertation Defense ____

Accepted by the Graduate School

____

Linda L. Black, Ed.D.
Associate Provost and Dean
Graduate School and International Admissions
Research and Sponsored Projects

ABSTRACT

Morgan, Michelle. *Standards-Based Grading Practices in Middle School Mathematics Classrooms: A Multicase Study*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2019.

This qualitative, multicase study sought to describe middle school mathematics teachers' teaching practices as they implemented standards-based grading. Specifically, the study focused on understanding middle school mathematics teachers' implementations of standards-based grading, use of assessment and feedback, and instructional design. Guided by cultural-historical activity theory lens, the researcher invited four middle school mathematics teachers who self-reported using standards-based grading practices to participate in the research study. Data collection for each case consisted of two interviews, lesson summaries and reflections, and classroom observations over the course of five consecutive class periods. Data analysis highlighted differences in the teachers' uses of mathematical tasks during instruction, implementation of instructional types, and teacher moves used to engage students in supporting student reasoning, assessment strategies, and evaluation practices. The evidence suggests the need for improved standards documentation, resource development, and professional development both at the preservice and inservice levels to better achieve the recommendations of the standards-based grading literature.

ACKNOWLEDGEMENTS

In pursuit of this degree, several individuals contributed to my eventual success. Their support and encouragement were instrumental to developing and maintaining the drive required to finish. First, I want to thank my Dad for fostering a passion for learning, a desire to strive for excellence, and a belief that I am capable of success. His interest in my work and belief in my abilities as an academic and educator pushed me to continue to better myself and my practice. His stories and advice with respect to teaching contributed significantly to my own teaching practice. Second, I want to thank Dr. Jeffrey King for keeping me focused when I needed focusing, reminding me to take occasional breaks, and reassuring me that there was an eventual end to this adventure. His willingness to debate the merits of standards-based grading helped me develop and refine my own beliefs and contributed to the completion of this study. Next, I want to thank my advisor, Dr. Robert Powers. Beginning with my Master's degree, he pushed me to be reflective in all areas of my practice. His willingness to discuss, question, and reflect on my work has helped me become a better educator and researcher. Finally, I want to thank my committee members, Drs. Steven Leth, William Blubaugh, and Mark Smith. Their practice of asking tough questions, and offering guidance and encouragement, pushed me to strive for the highest quality of work both in research and in my teaching practice.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In recent years, there has been movement towards implementing an alternative to traditional grading practices called standards-based grading (e.g., Gentile & Lalley, 2003; Vatterott, 2015). The foundation of this movement was the argument that historical perspectives on grading create uninterpretable grades and decrease motivation (Brookhart, 1994; Austin & McCann, 1992). However, advocates for an alternative grading practice argued that such a change requires more than simply changing the way in which teachers evaluate students (Vatterott, 2015). Currently, there is a lack of qualitative research focused on its implementation in the classroom (Brodersen & Randel, 2017). This study sought to fill that gap by describing the teaching practices of middle school mathematics teachers as they implemented standards-based grading.

Historical Perspectives On Grading

Early last century, Starch and Elliott (1912, 1913) raised significant concerns about the reliability of the percentage-score grading systems public schools used to measure student performance and academic accomplishments. In 1912, they found sizable amounts of error, up to 35 percentage points, when they asked multiple teachers to grade the same English exams for the same students as well as across multiple students. In a follow-up study, they found similar results with respect to grading and evaluating mathematics exams (Starch & Elliott, 1913). Both research studies contributed to a movement away from the percentage-score grading system that prevailed in the early 1900s to the letter-based grading system that is most prevalent today (Brookhart, 2012).

In the 1920s, educators hailed the adoption of a five letter-grade system (i.e., A, B, C, D, and F) as they perceived it as a fairer and more efficient method for reporting student performance and for sorting students into groups based on academic ability (Guskey, 1996; Vatterott, 2015). However, the use of fewer reported grades "served to reduce the variation of grades, but did not solve the problem of teacher subjectivity" (Guskey, 1996, p. 15). As a result, Crooks (1933) and others developed methods to adjust student grades to correct for bias in measurement. These methods have since become known as *grading on the curve* (Guskey, 1996). The basis for these methods was "the belief that among a sufficient number of students in school or college there [was] this so-called normal distribution of ability" (Crooks, 1933, p. 264). Specifically, there was evidence that scores of intelligence are normally distributed (Thorndike & Bregman, 1924) and many believed that teachers should directly link to and match such a distribution (Crooks, 1933). However, there was some disagreement about the correct distribution of grades with Corey (1930) arguing for an 8-24-36-24-8 distribution of grades and Davis (1930) and Eells (1930) arguing for a 6-22-44-22-6. By the early 1930s, Middleton (1933) claimed that American educators had come to an agreement that this system of grading on a curve was fair to all students and simple to administer.

A Modern View Of The Traditional System

The five-letter, curved grading system continued to be a topic of debate and frustration (Guskey, 1996; Vatterott, 2015). While controversial, this system became the primary method for measuring academic achievement and performance (Vatterott, 2015).

Critics of the five-letter grading system claimed that it creates meaningless grades (Deddeh, Main, & Fulkerson, 2010; Vatterott, 2015) which cause confusion that negatively impacts student motivation (Brookhart, 1994). Specifically, Vatterott (2015) argued that under a traditional grading system grades no longer "reflect proficiency in learning at all" (p. 19) when academic grades include measures of work ethic (e.g., penalties for late work) and motivational incentives (e.g., extra credit). According to Deddeh et al. (2010), "traditional grading practices often lead to 'grade fog,' in which the level of content mastery is distorted by such non-standards-based criteria as practice, neatness, organization, attendance, and behavior" (p. 54). That is, by adding *points* to a student's final grade that are unearned by mastering course content knowledge, teachers ultimately calculate a final grade that is difficult to use and interpret (Deddeh et al., 2010). Brookhart (1994) further argued that teacher use of hodgepodge grading leads directly to motivational issues.

Hodgepodge Grading

Contributing to issues of validity and interpretation, traditional grading philosophies and policies, sometimes called "hodgepodge grading," were typically inconsistent from student to student within teacher (Brookhart, 1991) as well as within and between school districts (Austin & McCann, 1992). Randall and Engelhard (2010) argued that issues with consistency are especially problematic in so called "borderline cases," because teachers must decide on how to grade students who fall between grades on the traditional letter grade system. Some argued that teachers' desires to adjust grades in these cases are based on their awareness "that even sound measures of achievement are imprecise and contain some error and, as such, [give] students the benefit of the doubt"

(Randall & Engelhard, 2010, p. 1378). However, Randall and Engelhard (2010) found that teachers were less likely to adjust students' grades upward in cases where students exhibited poor behavior or limited motivation. That is, only good students were given the benefit of the doubt. In addition, Guskey (2011) found that students' grades from the beginning of the term were an accurate predictor of final summative grades. The predictability of summative grades raises concerns about the impact and influence those early grades have on student achievement and motivation.

Motivational Concerns

Proponents of standards-based grading argued that traditional grading systems create an atmosphere where students are competing for the highest grade instead of competing for who learned or mastered more content. Specifically, Vatterott (2015) claimed that, under a traditional system, grades have become "the be-all and end-all, the goal itself, not an indicator of achieving the goal of learning" (p. 18). Some argued that grade-driven competition drives students to succeed in learning the material (Iamarino, 2014). However, as Iamarino (2014) noted, "in an environment that prioritizes points, students are often quick to identify and isolate the quickest methods of attaining those points, regardless of whether or not the activities they complete to get them are actually beneficial to the learning process" (p. 5). For example, in a study of middle school students, Anderman, Griesinger, and Westerfield (1998) found that students who described their educational experiences as performance-goal oriented reported increased the likelihood of cheating.

Standards-Based Grading Systems

There has been a movement in recent years towards adopting a more valid and reliable form of measurement and evaluation of academic achievement and performance (e.g., Gentile & Lalley, 2003; Vatterott, 2015). While not a new idea (e.g., Bloom, 1968), many stakeholders looked to standards-based grading practices for the answer (McMillan, 2009). According to McMillan (2009), "the fundamental purpose of standards-based grading is to compare student performance to established levels of proficiency in knowledge, understand, and skills" (p. 108). Variations to standards-based grading existed, including mastery-based grading (Gentile & Lalley, 2003), competency-based grading (Ryan & Cox, 2017), and specifications-based grading (Nilson, 2015), each of these systems share the same criterion-referenced philosophy. The present study focused on standards-based grading as that is how the school districts and teachers referred to their grading practices.

More Than A Grading System

While traditional grading systems determine a student's final grade by averaging or summing up grades earned over the course of a learning process, standards-based grading systems focus on the student's level of understanding at the end of the learning process (Iamarino, 2014). By using a standards-based grading system, Iamarino (2014) argued that a teacher "is better able to determine a student's grade based on the single most important aspect of education – how well the student comprehends the content of the course" (p. 2). Furthermore, standards-based grades better reflected actual student performance on district standardized exams (Deddeh et al., 2010).

In addition to how grades are determined, Vatterott (2015) claimed that standards-based grading systems differed from traditional grading systems in three more ways: how learning is (1) defined, (2) structured, and (3) experienced. First, in a standards-based grading system, learning is determined by the certain set of standards or goals that guide classroom instruction. However, instead of focusing on those standards superficially, learning requires a deeper level of understanding of the content described by those standards. As Vatterott (2015) noted, one measures rigor not by the number of concepts covered in a class, but rather by the level of mastery and higher-level thinking skills students acquire.

Second, in a standards-based grading system, learning is structured differently than in traditional settings. Within a traditional grading system, classroom instruction tends to follow a similar pattern: teach, test, and move on (Vatterott, 2015). However, in standards-based classrooms, teachers present learning opportunities in a spiral format (Gentile & Lalley, 2003; Vatterott, 2015). A typical pattern for spiraling the curriculum is to "teach, check for understanding, apply learning, get feedback, revise learning, and get more feedback until mastery is achieved" (Vatterott, 2015, p. 29). In addition, teachers differentiate learning to meet the needs of individual students. When differentiating content, teachers provide access to material based on the student's current understanding with the goal of helping all students achieve their maximum potential (Tomlinson & Eidson, 2003).

Third, learning is experienced differently in a standards-based grading system than in traditional settings. With standards-based grading systems, educators consider learning as a process that happens over the course of many attempts at success. The belief is that teachers should not punish students for failing to achieve mastery or proficiency while they are learning a concept (Deddeh et al., 2010). This means that teachers record only the final attempt at demonstrating mastery in the gradebook. According to Vatterott (2015), “when we release students from the stigma of failure and when we use feedback instead of grades during the process of learning, students soon develop perseverance based on the expectation of success” (p. 33). That is, students become intrinsically motivated to learn the material instead of extrinsically motivated to earn a grade (Deddeh et al., 2010).

Support From The Literature

Beyond clearing up confusion regarding the interpretation of student grades, there was evidence in the literature that standards-based grading systems better predict success on standardized exams when compared to traditional grading systems (Pollio & Hochbein, 2015), improve student motivation (Vatterott, 2015), and provide better information for instructional planning (Guskey, Swan, & Jung, 2010).

Correlation with state-assessments. Deddeh et al. (2010), Bradbury-Bailey (2011), and Hochbein and Pollio (2016) all reported significant correlations between students’ standards-based grades and state-standardized assessment results. Those correlations were strongest when considering the results of minority students (Bradbury-Bailey, 2011; Hochbein & Pollio, 2016). Furthermore, students who experienced standards-based grading practices scored higher on the state-standardized assessment when compared to their peers who experienced traditional grading practices (Hochbein & Pollio, 2016).

Fosters growth mindset. Franklin (2016) found that Grade 7 "students exposed to a standards-based grading model reported a higher frequency of growth mindset characteristics in the areas of effort in math and goal setting compared to their peers exposed to a traditional grading model" (p. 87). Specifically, when teachers allowed students to redo and retake assessments, given frequent opportunities for feedback, and given multiple chances to demonstrate their knowledge, students were better able "to observe their own growth, thus continuing to build a belief in their ability to increase their intelligence" (p. 100). As a result, the students' motivation to learn improved.

Better informed instruction. Guskey et al. (2010) found that teachers believed that standards-based grading provided more useful information when compared to traditional grading systems. For example, standards-based grading allowed teachers to identify specifically what content students had mastered and which content they still need instruction. When using standards-based grading, Teachers also reported being able to focus more on deepening their students' understanding of content rather than attempting to cover a breadth of knowledge (Hochbein & Pollio, 2016). While many teachers noted that standards-based grading was more time consuming, they reported believing it was worth the effort In The Long Run (Guskey et al., 2010; Hochbein & Pollio, 2016).

Issues With Implementation

Gentile and Lalley (2003) noted that, while standards-based grading is more than a modification to the course grading scheme, many teachers fail to appropriately modify all aspects of the learning process. Furthermore, Gentile and Lalley (2003) argued that many of those who claim to be utilizing a standards-based grading philosophy make several common errors, including (1) ending the learning cycle once a student reaches proficiency, (2) applying a traditional (norm-referenced) grading system within the nominal structures of standards-based grades, and (3) focusing on lower-level knowledge and recall tasks when assessing proficiency.

In a study of one Colorado school district, Brodersen and Randel (2017) found that "teachers may have promoted students too quickly by scoring students as competent on their learning targets before they truly were competent" (p. 11). In this case, their evidence suggested that

> When students passed a district assessment, teachers entered into the learning management system scores designating students as proficient for all learning targets, rather than entering scores on an ongoing basis as the gathered proficiency information relevant for each individual learning target. (p. 12)

According to Iamarino (2014) and Welsh and D'Agostino (2009), implementation of reform-grading systems can be difficult and uncomfortable for teachers as well as other stakeholders such as district supervisors, students, and parents. Specifically, teachers were reluctant to give lower grades to students out of fear that it would negatively impact student motivation (Hochbein & Pollio, 2016). In addition, many teachers raised concerns about failing to reinforce positive behaviors and work habits through the use of grades (Tierney, Simon, & Charland, 2011). With respect to all stakeholders, many agreed that traditional, hodgepodge grading resulted in issues relating to validity and interpretation; however, "that is what they expect and endorse" (Cross & Frary, 1999, p. 70). That is, hodgepodge grading has become so engrained in the education system that students, teachers, administrators, and parents are willing to accept bias in grades.

Need For Further Research

Current research into standards-based grading practices have been largely quantitative in nature focused on the alignment between standards-based grades and standardized assessment scores (Bradbury-Bailey, 2011; Deddeh et al., 2010; Hochbein & Pollio, 2016) or a survey of teachers' grading practices (Hochbein and Pollio, 2016; Tierney et al., 2011). Brodersen and Randel (2017) and Pollio and Hochein (2015) noted a lack of qualitative research focused on classroom implementation of standards-based grading practices and its impact of classroom instruction. This multicase study sought to fill this gap in the literature by exploring the impact of standards-based grading systems on the practices of middle school mathematics teachers.

Cultural-Historical Activity Theory

This research study was guided by cultural-historical activity theory (Engeström, 2015). Cultural-historical activity theory is a perspective of human cognition which takes as its minimal unit an activity system which consists of six components: subject, object, instrument, community, division of labor, and rules (Engeström & Sannino, 2010). Table 1 provides a brief description of each component. According to Engeström (2015), each component only takes on meaning within the larger activity system. Analyzing the interconnected components of the activity system (see Figure 1) allows researchers "to explain change, learning, and development as an *immanent* feature of a system rather than in terms of externally produced cause-effect relations" (Roth, 2014, p. 11, emphasis in original).

Table 1

Definitions of activity system components

Component	Definition
Subject	The individual or subgroup whose position and point of view are chosen at the perspective of the analysis.
Object	The 'raw material' or 'problem space' at which the activity is directed. The object is turned into outcomes.
Instrument	The tools and signs used.
Community	The individuals and subgroups who share the same general object.
Division of Labor	The horizontal division of tasks and vertical division of power and status.
Rules	The explicit and implicit regulations, norms, conventions, and standards that constrain actions.

Note. Adapted from Engeström and Sannino (2010, p. 6).

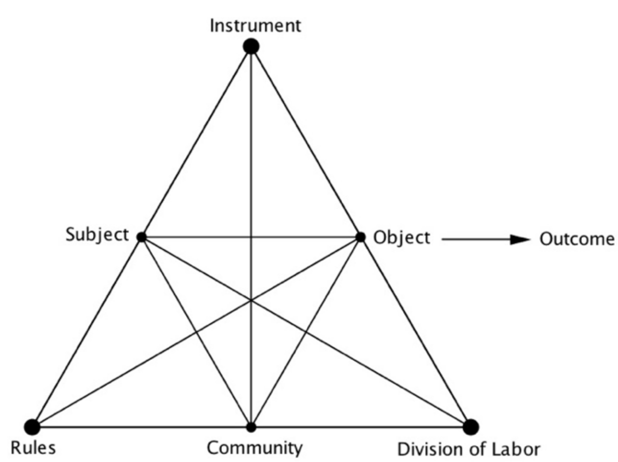

Figure 1. The structure of human activity (adapted from Engeström, 2015, p. 63).

Theory Of Expansive Learning

With respect to cultural-historical activity theory, learning and collective change occurs as the result of "contradictions [or tensions] within and between activity systems" (Engeström, 2000, p. 309). Specifically, "as more actors join in, a collaborative analysis and modeling of the zone of proximal development are initiated and carried out" which results in the development of a different model of the activity system (Engeström & Sannino, 2010, p. 5). Engeström and Sannino (2010) called this process the theory of expansive learning.

According to Engeström and Sannino (2010), the theory of expansive learning relies on the metaphor of expansion whereby "learners learn something that is not yet there" (p. 2). That is, "the learners construct a new object and concept for their collective activity, and implement this new object and concept in practice" (Engeström & Sannino, 2010, p. 2). Engeström (2000) noted that learners construct such knowledge from both a developmental perspective and a collective perspective. Expansive learning is the process of developing and resolving "successively evolving contradictions" (Engeström & Sannino, 2010, p. 7). According to Engeström and Sannino (2010), conflict experiences are those tensions which occur as the result of short-term action as opposed to developmentally significant contradictions which occur on the level of activity and over a much longer period. They claimed that such contradictions occur in one of four ways:

> (a) as emerging latent primary contradictions within each and any of the nodes of the activity system, (b) as openly manifest secondary contradictions between two or more nodes (e.g., between a new object and an old tool), (c) as tertiary contradictions between a newly established mode of activity and remnants of the previous mode of activity, or (d) as external quaternary contradictions between the newly reorganized activity and its neighboring activity systems. (p. 7)

Activity Theory In The Classroom

One of the ways to document changes in classroom activity during the instruction of mathematics lessons is through the lens of cultural-historical activity theory (e.g., Tomaz & David, 2015). Herbst and Chazan (2003) argued that several factors shape school mathematical activity. On one hand, they claimed that the mathematics teacher's personal and professional goals play a significant role in shaping the approach and concepts focused on during instruction. On the other hand, they acknowledged that the teacher might choose approaches or focus on certain concepts as the result of outside influences including school polices and student learning needs. Therefore, to understand classroom instruction, it is important to consider all components of the teacher's activity system (see Table 2). That is, from the teacher's perspective (i.e., the subject's perspective), it is important to consider the mathematical goals of the lesson (i.e., the object), and the lesson plan and other instructional materials (i.e., the instruments) available during instruction (Herbst & Chazan, 2003). In addition, it is important to consider the interactions (i.e., the division of labor) of the teacher, students, and other stakeholders (i.e., the community) with respect to expectations and classroom norms (i.e., the rules) (Herbst & Chazan, 2003).

Table 2

Cultural-historical activity theory component examples for a mathematics classroom

Component	Mathematical Classroom
Subject	Mathematics Teacher
Object	Mathematical goals of the lesson
Instrument	Lesson plan, classroom materials, manipulatives, discourse, etc.
Community	Teacher, students, school, district, parents, etc.
Division of Labor	Professional and personal obligations to the content, students, district, etc.
Rules	Commitments, expectations, beliefs, classroom norms, etc.

Note. *Adapted from Herbst and Chazan (2003).*

In order to analyze the six components of classroom instructional activity, Wells (1996) identified two levels of instruction: (1) macro-level and (2) micro-level. At the macro-level, the teacher is responsible for developing an appropriately challenging lesson that attends to the goals and expectations associated with the activity system's rules, community, and division of labor. That is, at the macro-level, the teacher must create a lesson that attends to components of the activity system that might extend beyond the immediate lesson. At the micro-level, the teacher is responsible for ensuring that they implement the lesson as planned. Specifically, the micro-level of instruction focuses on the "moment-by-moment co-construction of meaning" when "the teacher observes how students take it up, both individually and collectively, and acts to assist them in whatever way seems most appropriate to enable them to achieve the goals that have been negotiated" (Wells, 1996, p. 83). The teacher must attend to themself as the subject of the activity system, the goals or objects of the lesson, and the instrumentation used during instruction.

During the implementation of instruction, contradictions and tensions inevitably arise from the perspective of the teacher. The teacher must simultaneously navigate the macro- and micro-levels of teaching while adjusting the flow of the lesson in order to reconcile these contradictions and tensions in order to achieve the goals of the lesson. Wells (1996) noted that

> The subject-object relationship – that is to say, the subject's goal orientation – is modified by the cultural rules that apply to this relationship and by the division of labor in which it is embedded. These rules, or norms, might well include the tools considered appropriate to use, and the way in which control of their use is distributed among the different categories of community members who are regularly involved in this and related 'actions.' However, these relationships are not static; they are

> continuously being constructed and reformulated in the course of their deployment in particular situated 'actions'. (p. 76).

Classroom activity and instructional moves are the result of the teacher's responses to perceived contradictions and tensions. That is, the implementation of a lesson is the result of the teacher's goal-directed behavior called actions (Wells, 1996).

Research Study Overview

The purpose of this qualitative, multicase study was to describe middle school mathematics teachers' teaching practices as they implemented stands-based grading. Specifically, the study sought to answer the following research questions:

Q1 How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?

Q1a What is the nature of the mathematical tasks selected by middle school mathematics teachers for instruction?

Q1b How do middle school mathematics teachers facilitate mathematical instruction?

Q1c What is the nature of classroom discourse as facilitated by middle school mathematics teachers during instruction?

Q1d How do middle school mathematics teachers utilize assessment strategies?

Q2 How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?

Q3 What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?

Q3a How do middle school mathematics teachers' own teaching philosophies impact their implementation of standards-based grading practices?

Q3b How do other stakeholders (e.g., school district personal, parents, students) impact middle school mathematics teachers' implementation of standards-based grading practices?

After a review of the literature (see Chapter II), the researcher invited four middle school mathematics teachers to participate in data collection (see Chapter III). Following data analysis, the four case study reports (see Chapters IV through VII) and cross-case analysis (see Chapter VIII) identified similarities and differences among the four teachers with respect to their activity systems and subsequent implementation of standards-based grading. With consideration of the evidence from that data analyses, the researcher argued for increased professional development for preservice and inservice teachers as well as resource development for teachers implementing standards-based grading (see Chapter IX).

CHAPTER II

LITERATURE REVIEW

Criticisms of traditional grading policies and calls for reform are not new (Guskey, 2009). Guskey (2009) cited Bloom (1968) as one of the first to argue against traditional grading policies in favor of polices that support *mastery* learning. In 1968, Benjamin Bloom argued that “most students (perhaps over 90 percent) can master what we have to teach them, and it is the task of instruction to find the means which will enable out students master the subject under consideration” (Bloom, 1968, p. 1). Bloom (1968) believed that to develop such a strategy we must have a better understanding of the individual needs and differences of our students and the resulting impact on the instructional process. To achieve such a strategy, it would require a significant shift in the attitudes of educational stakeholders as well as the uses of evaluation and assessment (Bloom, 1968).

In a survey of Colorado mathematics teachers, Morgan and Powers (2018) found that several school districts in Colorado are mandating that their teachers use standards-based grading practices as part of their classroom instruction. Heflebower, Hoegh, and Warrick (2014) argued that, while some practitioners view this change as a change in grading, successful implementation of standards-based grading requires a complete shift in teaching practices (see also Gentile & Lalley, 2003; Vatterott, 2015). For example, Gentile and Lalley (2003) identified six elements of mastery learning that many teachers using a standards-based grading system tend to neglect: (1) lessons need to be designed in a spiral curriculum; (2) prerequisite knowledge needs to be measured prior to the start of lesson so that issues and misconceptions can be addressed; (3) instruction employs multiple teaching strategies which foster high-level thinking; (4) objectives for mastery need to be clearly articulated to students; (5) teachers must use a variety of assessments to measure for mastery; and (6) remediation is available and planned for in order to help all students achieve mastery. Therefore, according to Vatterott (2015), beyond how teachers use and calculate grades, standards-based grading systems differ from traditional grading systems in how educators define and structure learning opportunities as well as how students experience those opportunities.

How Teachers Use Grades: Focus On Learning

In 1983, the National Commission on Excellence in Education argued that "grades should be indicators of academic achievement so they can be relied on as evidence of a student's readiness for further study" (p. 27). According to Guskey (2009), many stakeholders agreed "that grades should reflect how well students have achieved the learning goals established for a subject area or course" (p. 18). However, how teachers calculate those grades differed from teacher to teacher. Guskey (2009) defined three categories of grades: (1) process, (2) progress, and (3) product. Process and progress criteria both included the idea that "grades should reflect not only the final results but also *how* students got there" (Guskey, 2009, p.18, emphasis in original). That is, these criteria foster ideas associated with traditional grading practices. Product criteria, however, focused "on *what* students know and are able to do at a particular point in time" (Guskey, 2009, p. 18, emphasis in original). As a result, many advocates for standards-based grading argued that product criteria should guide the grading process

(Guskey, 2009; Heflebower et al., 2014). Specifically, advocates for standards-based grading argued that grades should only reflect students' current understanding of content-specific knowledge and students should have several opportunities to demonstrate that knowledge (Heflebower et al., 2014; Vatterott, 2015).

Grades That Reflect Learning

To have grades that reflect students' understanding of content-specific knowledge, teachers should organize learning opportunities around learning targets or standards which detail the specific knowledge students need to obtain (Vatterott, 2015). Instead of creating a single grade based on some combination of those targets, advocates recommended that teachers give students an individual grade for each learning target or standard and not include measurements of nonacademic behavior in student grades (Heflebower et al., 2014; Gentile & Lalley, 2003; Vatterott, 2015). According to Vatterott (2015), this grading scheme should motivate students to continue to work on obtaining the knowledge necessary for success on each learning target or standard.

Multiple Opportunity Grades

Advocates of standards-based grading argued that, if students' grades are supposed to reflect their current understanding of content-specific material, then those grades should only include measurements of students' most recent attempt to demonstrate their understanding (Heflebower et al., 2014; Vatterott, 2015). That is, "students are not penalized with grades while they are still learning" (Vatterott, 2015, p. 36). The recommendation was that the teacher should replace old measurements of student understanding when they collect new measurement data (Vatterott, 2015). Furthermore, the advice was that the teacher should give students multiple opportunities to demonstrate their knowledge and not penalize students for the number or timing of those opportunities (Heflebower et al., 2014; Vatterott, 2015).

How Teachers Define Learning: Quality Over Quantity

Critics of traditional grading practices claimed that such practices focus too much attention on rote memorization and procedural skills (Gentile & Lalley, 2003; Vatterott, 2015). In contrast, advocates of standards-based grading argued that rigor should not be defined by the quantity of knowledge obtained, but rather "by the complexity of tasks and the level of mastery of higher-level thinking skills that students can attain" (Vatterott, 2015, p. 28). That is, "learning is defined by the standards – not by what students *know*, but by what they can *do with what they know*" (Vatterott, 2015, p. 27, emphasis in original). As a result, the recommendation was that teachers need to make sure that their instruction aligns with clearly defined academic standards and address high-cognitive demand tasks (Smith & Stein, 1998).

Academic Standards

Bigham (2015) defined standards as "clearly defined statements of the knowledge and skills students should have at each grade that prepares them for the next grade" (p. 4). The belief was that well-defined academic standards offer numerous benefits including clear definitions of what students should know, common goals and greater equity for all classrooms and students, and consistent communication about student achievement (O'Connor, 2009). In the United States, the development and implementation of academic standards was met with resistance and controversy (Bigham, 2015).

National standards movement. When it comes to academic standards, a common misconception was that they tell teachers exactly how students should learn the specific content defined in those standards (Bigham, 2015). Bigham (2015) argued that this is not the case. Teachers assist students in achieving standards with the help of a curriculum (Bigham, 2015). That is, "the textbooks, materials, instructional techniques, and other resources use to teach standards" (Bigham, 2015, p. 5). Because of this misconception, movement towards creating a national list of academic standards was met with resistance (Bigham, 2015).

In 1983, the National Commission on Excellence in Education released the *A Nation at Risk* report which suggested that the quality of K-12 education in the United States was declining. Among their recommendations, the commission recommended that schools "adopt more rigorous and measurable standards, and higher expectations, for academic performance and student conduct" (National Commission on Excellence in Education, 1983, p. 27). Many viewed attempts at creating such standards as federal interference in state-run education with federal representatives attempting to dictate how teachers should teach students across the country (Bigham, 2015). Many critics of this effort cited the *Elementary and Secondary Education Act of 1965* which forbade "the federal government from mandating to states how to teach, what to teach, or what resources to use in instruction at the state and local level" (Bigham, 2015, p. 14).

In 1994, with the reauthorization of the *Elementary and Secondary Education Act of 1965*, called the *Improving America's Schools Act of 1994*, all states were mandated to develop, at a minimum, academic standards in mathematics and language arts. Specifically, these standards were to include:

> (i) Challenging content standards in academic subjects that – (I) specify what children are expected to know and be able to do; (II) contain coherent and rigorous content; and (III) encourage the teaching of advanced skills; (ii) challenging student performance standards that – (I)

> are aligned with the State's content standards; (II) describe two levels of high performance, proficient and advanced, that determine how well children are mastering the material in the State content standards; and (III) describe a third level of performance, partially proficient, to provide complete information about the progress of the lower performing children toward achieving to the proficient and advanced levels of performance. (p. 7)

While policy experts anticipated states would create such standards, they did not expect, however, states to collaborate with each other or even report their standards to the U.S. Secretary of Education. As a result, the level of rigor, quality, and focus of the standards varied greatly from state to state (Bigham, 2015). In addition to developing state academic standards, the *Improving America's Schools Act of 1994* required that states development assessments to measure those standards to measure student achievement and to hold schools accountable; that is, to measure adequate yearly progress in schools.

In 2001, with the passing of the *No Child Left Behind Act of 2001*, another reauthorization of the *Elementary and Secondary Education Act of 1965*, the federal government required schools to create annual, public reports of their assessment data and the act penalized states for failing to meet measures of annual progress. During the years that followed, the public reports highlighted "pronounced discrepancies between students' performance on state and national level assessments" (Bigham, 2015, p. 12). According to Bigham (2015), *No Child Left Behind* "galvanized states in a positive way in that they now had common goals and challenges that brought them to the discussion table" (p. 7). At this point, the state level drove the call for national academic standards as opposed to previous attempts from the federal level.

Common core state standards. In 2010, the National Governors Association Center for Best Practices [NGA] and Council of Chief State School Officers [CCSSO] released the *Common Core State Standards*. With this publication, the United States had a common, national set of academic standards for both mathematics and language arts. With respect to mathematics, the common core includes both grade-level, content-specific standards as well as across grade-level standards of mathematical practice. In contrast to previous state standards documents which emphasized basic skills and calculation, the common core state standards in mathematics increased emphasis on conceptual understanding, critical thinking, and communication of mathematics ideas (Bigham, 2015; NGA & CCSSO, 2010).

Colorado academic standards in mathematics. In June 2010, the Colorado Department of Education [CDE] conducted a gap analysis between the then current state standards and the common core state standards, in conjunction with an independent analysis, and found a 95% alignment between the standards (CDE, 2010). Later, in August 2010, the Colorado State Board of Education decided to adopt the common core state standards and initiated the process to integrate them into the existing state standards (CDE, 2010). The resulting document consists of near equivalent versions of the common core with the addition of elements related to “personal financial literacy, 21st century skills, school readiness competencies, postsecondary and workforce readiness competencies, and preschool expectations” (CDE, 2010, p. 1).

While the wording aligns closely with the standards outlined in the common core, the structure of the Colorado academic standards was different (CDE, 2010). The Colorado academic standards document address four content standards: (1) number sense, properties, and operations; (2) patterns, functions, and algebraic structures; (3) data analysis, statistics, and probability; and (4) shape, dimension, and geometric relationships. Each standard was then sub-divided into grade level expectations and evidence outcomes. For example, standard three for eighth grade mathematics consisted of three grade level expectations divided into 25 evidence outcomes. In total, teachers need to address 55 evidence outcomes to cover all the required content for eighth grade mathematics.

Implications for standards-based grading. When measuring student achievement and understanding, advocates for standards-based grading recommended that educators report results based on the individual student (Brookhart, 2012; Gentile & Lalley, 2003; Vatterott, 2015). To accomplish this, some teachers used the state standards as written while others rewrote those standards into student-friendly terms (Brookhart, 2012). In either case, successful implementation of standards-based grading required that students understand the standards (or learning targets) they must achieve and the measurement criteria used to determine that achievement (Vatterott, 2015).

High-Cognitive Demand Tasks

As noted above, in addition to including procedural fluency, the common core state standards in mathematics increased emphasis on conceptual understanding, critical thinking, and communication of mathematics ideas (Bigham, 2015; National Governor's Association Centers for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010). According to the common core state standards in mathematics, "mathematical understanding and procedural skill are equally important, and both are assessable using mathematical tasks of sufficient richness" (NGA & CCSSO, 2010, p. 4).

Stein, Grover, and Henningsen (1996) defined a mathematical task "as a classroom activity, the purpose of which is to focus students' attention on a particular mathematical idea" (p. 460). For each task, Doyle (1988) defined four elements: (a) goal or end product; (b) conditions and resources; (c) operations involved; and (d) the importance to the work system. Doyle (1983) argued that academic tasks can be categorized into four categories: (1) memory; (2) procedural/routine; (3) comprehension/understanding; and (4) opinion. See Table 3 for descriptions.

Table 3

Categories of academic tasks

Category	Description
Memory Tasks	Students are expected to recognize or reproduce information previously encountered.
Procedural or Routine Tasks	Students are expected to apply a standardized and predictable formula or algorithm to generate answers.
Comprehension or Understanding Tasks	Students are expected to (a) recognize transformed or paraphrased versions of information previously encountered, (b) apply procedures to new problems or decide from among several procedures those which are applicable to a particular problem, or (c) draw inferences from previously encountered information or procedures.
Opinion Tasks	Students are expected to state a preference for something.

Note. Adapted from Doyle (1983, pp. 162-163).

In addition to type of task, both Doyle (1988) and Stein et al. (1996) recommended comparing academic tasks based on cognitive demand. That is, the type and depth of thinking required of students when engaging with the task (Doyle, 1988; Stein et al., 1996). To determine cognitive demand, teachers needed to “consider the students – their age, grade level, prior knowledge and experiences – and the norms and expectations for work in their classroom” (Smith & Stein, 1998, p. 344). Smith and Stein (1998) defined four levels of cognitive demand related to mathematical tasks (see Table 4 for low-level tasks and Table 5 for high-level tasks). Stein et al. (1996) characterized high-cognitive demand tasks as those mathematical tasks that require multiple solution strategies, allow for multiple representations, or require students to explain and justify their reasoning.

Table 4

Low-levels of cognitive demand for mathematical tasks

Category	Description
Memorization	• Involve either reproducing previously learning facts, rules, formulas, or definitions or committing facts, rules, formulas or definitions to memory. • Cannot be solved using procedures because a procedure does not exist or because the time frame in which the task is being completed is too short to use a procedure. • Are not ambiguous. Such tasks involve the exact reproduction of previously seen material, and what is to be reproduced is clearly and directly stated. • Have no connection to the concepts or meaning that underlie the facts, rules, formulas, or definitions being learned or reproduced
Procedures Without Connections	• Are algorithmic. Use of the procedure either is specifically called for or is evident from prior instruction, experience, or placement of the task. • Require limited cognitive demand for successful completion. Little ambiguity exits about what needs to be done and how to do it. • Have no connection to the concepts or meaning that underlie the procedure being used. • Are focused on producing correct answers instead of on developing mathematical understanding. • Require no explanations or explanations that focus solely on describing the procedure that was used.

Note. From Smith and Stein (1998, p. 348).

Table 5

High-levels of cognitive demand for mathematical tasks

Category	Description
Procedures With Connections	• Focus students' attention on the use of procedures for the purpose of developing deeper levels of understanding of mathematical concepts and ideas. • Suggest explicitly or implicitly pathways to follow that are broad general procedures that have close connections to underlying conceptual ideas as opposed to narrow algorithms that are opaque with respect to underlying concepts. • Usually are represented in multiple ways, such as visual diagrams, manipulatives, symbols, and problem situations. Making connections among multiple representations helps develop meaning. • Require some degree of cognitive effort. Although general procedures may be followed, they cannot be followed mindlessly. Students need to engage with conceptual ideas that underlie the procedures to complete the task successfully and that develop understanding.
Doing Mathematics	• Require complex and non-algorithmic thinking – a predictable, well-rehearsed approach or pathway is not explicitly suggested by the task, task instructions, or a worked-out example. • Require students to explore and understand the nature of mathematical concepts, processes, or relationships. • Demand self-monitoring or self-regulation of one's own cognitive processes. • Require students to access relevant knowledge experiences and make appropriate use of them in working through the task. • Require students to analyze the task and actively examine task constraints that may limit possible solution strategies and solutions. • Require considerable cognitive effort and may involve some level of anxiety for the student because of the unpredictable nature of the solution process required.

Note. From Smith and Stein (1998, p. 348).

Over the course of implementation and instruction, tasks can change form and function (Stein et al., 1996). See Figure 2 for a representation of task-related variables and student learning outcomes. Specifically, Stein et al. (1996) found that teachers modified tasks when they incorporate them into their lesson plans as well as when they implement them during instruction. While in many cases teachers decreased the cognitive demand of tasks (Stein et al., 1996), Ellis, Ozgur, and Reiten (2019) argued that teachers can support students in mathematical reasoning by utilizing certain instructional moves. Ellis and colleagues identified teachers moves that have low and high potential for (a) eliciting, (b) responding to, (c) facilitating, and (d) extending student reasoning. See Figure 3 for their teacher moves for supporting student reasoning [TMSSR] framework. By utilizing high-level moves, allowed teachers to "focus on the students' ideas, enabling teachers to provide students with a space to engage meaningfully in the processes of mathematical reasoning" (Ellis et al., 2019, p. 21). That is, teachers were better able to maintain a high level of cognitive demand (Stein et al, 1996).

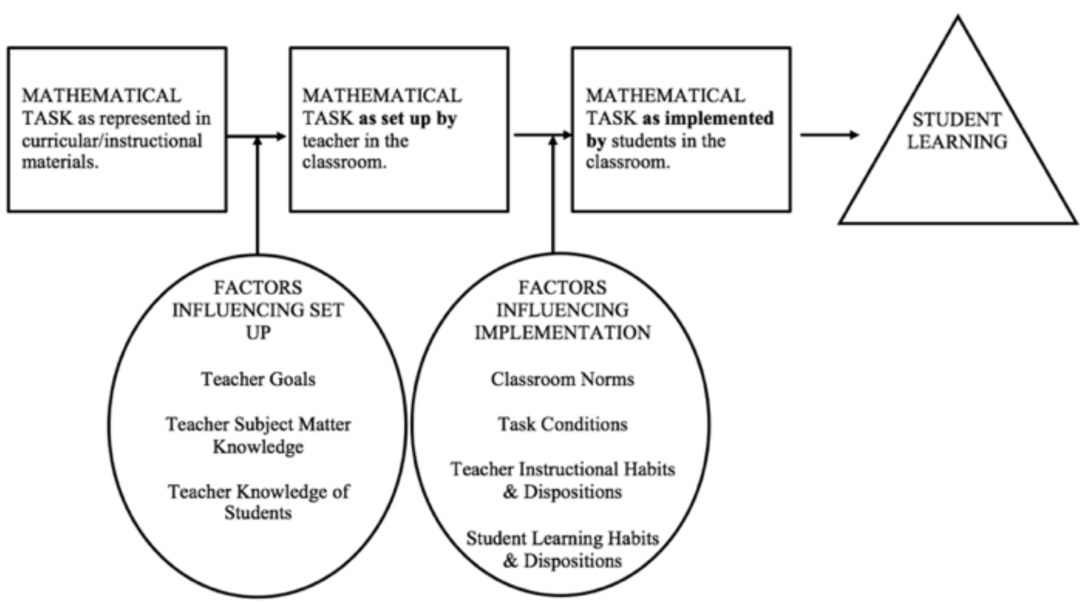

Figure 2. Relationship among various task-related variables and student learning (Stein et al., 1996, p. 459).

Eliciting Student Reasoning		Responding to Student Reasoning	
Low ←	→ High	Low ←	→ High
Eliciting Answer	Eliciting Ideas	Correcting Student Error	Prompting Error Correction
Eliciting Facts or Procedures	Eliciting Understanding	Re-voicing	Re-Representing
Asking for Clarification	Pressing for Explanation	Encouraging Student Re-voicing	
Figuring Out Student Reasoning		Validating a Correct Answer	
Checking for Understanding			

	Facilitating Student Reasoning		Extending Student Reasoning	
	Low ←	→ High	Low ←	→ High
Guiding	Cueing	Providing Guidance	Encouraging Evaluation	Encouraging Reflection
Guiding	Funneling	Encouraging Multiple Solution Strategies	Pressing for Precision	Encouraging Reasoning
Guiding	Topaze Effect	Building	Topaze for Justification	Pressing for Justification
Providing	Providing Information	Providing Alternative Solution Strategies		Pressing for Generalization
Providing	Providing Procedural Explanation	Providing Conceptual Explanation		
Providing	Providing Summary Explanation			

Figure 3. The teacher moves for supporting student reasoning framework (Ellis et al., 2019, p. 11).

According to Doyle and Carter (1984), when educators embedded academic tasks in an evaluation system, there existed a certain level of ambiguity and risk. Furthermore, teachers and students affected the nature of those tasks as they attempted to manage ambiguity and risk (Doyle & Carter, 1984),. For example, students sought "to reduce ambiguity and risk by clarifying task demands and obtaining feedback concerning the quality of their provisional writing efforts" (Doyle & Carter, 1984, p. 145). The conclusion was that teachers had to choose between maintaining classroom motivation and potentially reducing task ambiguity. Doyle and Carter (1984) suggested that teachers most often reduced ambiguity in favor of maintaining classroom order. Unfortunately, such decisions also typically resulted in a reduction of cognitive demand (Henningsen & Stein, 1997). For example, by providing too much guidance or information during the implementation of the task, the teacher reduced the task ambiguity by providing students the answers to more complex components of the task (Henningsen & Stein, 1997). Henningsen and Stein (1997) claimed that this pedagogical approach results in the students focusing solely on procedures without any attention on the underlying concepts or meaning the task intends to address.

How Teachers Structure Learning Opportunities: A Formative Spiral

According to Vatterott (2015), traditional grading systems have a similar instructional structure: teach, test, and move on. Under these systems, some argued that all students should receive the same instruction and learning opportunities (Gentile & Lalley, 2003; Heflebower et al., 2014; Vatterott, 2015). After a fixed amount of time, all students received the same test whose grade was a permanent addition to a student's academic record (Vatterott, 2015). Under such a system, Boaler (2016) and Vatterott (2015) argued that student achievement varies with some students achieving high-levels of understanding and other students failing to achieve.

In contrast, advocates for standards-based grading argued that instruction should adjust to the individual needs of each student based on the results of continuous formative assessment and feedback (Tomlinson, 2014a; Vatterott, 2015). Vatterott (2015) argued that this practice reverses the learning and achievement relationship. That is, when teachers varied the amount of time they gave students to learn, then they fixed student achievement at a high level (Vatterott, 2015). The aim of standards-based grading was to give all students the resources, including time, necessary to achieve a high-level of understanding. To achieve this aim, the structure of learning opportunities should include continuous formative assessment, differentiated instruction, effective feedback, and multiple opportunities to complete summative assessments (Gentile & Lalley, 2003; Heflebower et al., 2014; Vatterott, 2015).

Continuous Formative Assessment

A key characteristic of the structure of learning opportunities as part of a teacher's implementation of a standards-based grading philosophy is the use of continuous formative assessment (Vatterott, 2015). Black and Wiliam (1998) defined assessment as "all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities" (p. 140). When teachers used such information to improve teaching and learning opportunities during subsequent lessons, then those assessments were called formative (Black & Wiliam, 1998; Chappuis, Stiggins, Chappuis, & Arter, 2012). Specifically, Black and Wiliam (2009) argued that:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (p. 9)

Bloom (1968) advised that, when used appropriately, formative assessments "pace the learning of students and help motivate them to put forth necessary effort at the appropriate time" (p. 9). Given the focus on improving learning opportunities, many referred to formative assessment as assessment for learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Chappuis et al., 2012; Danielson, 2008; Stiggins, 2005; Vatterott, 2015).

Wiliam and Thompson (2008) identified five strategies for integrating formative assessment with improving students' opportunities to learn (see Figure 4). They categorize these strategies based on the questions the assessment information answers: (1) where the learner is going; (2) where the learner is right now; and (3) how to get there. To answer the first question, Wiliam and Thompson (2008) recommended making learning intentions and criteria for success clear to all members of the classroom (i.e., the teacher and students). The recommendation was that the teacher should share and clarify their intentions and criteria as well as provide an opportunity for students to discuss those intentions and criteria. From the perspective of the teacher, they recommended that teachers design classroom activities that elicit evidence of student learning as a strategy for answering the second question and providing effective feedback as a strategy for answering the third question. Finally, they recommended fostering peer-evaluation and self-evaluation as strategies for engaging the students in answering the second and third questions for themselves.

	Where the Learner Is Going	Where the Learner Is Right Now	How to Get There
Teacher	Clarifying and sharing learning intentions and criteria for success	Engineering effective classroom discussions and tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success.	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success.	Activating students as the owners of their own learning.	

Figure 4. Framework relating strategies of formative assessment to instructional processes (Wiliam & Thompson, 2008, p. 63).

Guskey (1997) argued that teachers often neglect formative assessment or implement it inappropriately. The rationale was that since the goal of formative assessment is to improve future teaching and learning opportunities, the purpose of these assessments should be to gather information about how students are progressing in understanding the content. As a result, advocates for standards-based grading argued that teachers should not use formative assessment as part of grade calculations or evaluation (Guskey, 1997; Tomlinson, 2014a; Vatterott, 2015); instead, such assessments should provide feedback to the teacher and students to adjust instruction and learning (Guskey, 1997; Popham, 2008). Tomlinson (2014a) argued by removing the fear of judgement from formative assessments, student responses are more likely to reflect the students' actual current level of understanding.

Popham (2008) identified four levels of formative assessment. At the first level, teachers use formative assessment collected from students to adjust their instructional approach. The second level is comprised of students using the evidence to adjust how they approach learning. At the third level, there is a cultural shift in the classroom climate of assessment from a way to compare and rank students to a way of transforming teaching and learning. Finally, the fourth level encompasses a systematic approach to formative assessment that occurs at the school level.

According to Popham (2008), there existed four steps for successful implementation of level one formative assessments (see Table 6). The first step of implementation occurred when the teacher identified adjustment occasions during their planned lesson. Popham (2008) defined these occasions as "the most significant choice-points associated with students' movement toward mastery of the target curricular aim" (p. 53). That is, adjustment occasions marked key moments during instruction where students should have obtained a certain level of understanding to achieve success in future learning opportunities. The second step of implementation occurred when the teacher planned for formal and informal assessments to use for each adjustment occasion. When the teacher identified, in advance of instruction, adjustment triggers was the third step of implementation (see Figure 5). That is, teachers must

> Establish, *before* collecting assessment evidence from students, (1) a minimum per-student performance level and (2) a minimum per-class performance level, that is, the percentage of students who much achieve

> the minimum per-student performance level. (Popham, 2008, p. 64, emphasis in original)

The final step of implementation occurred when the teacher made instructional adjustments during instruction and in the planning of future lessons.

Table 6

Steps for implementing level one formative assessments

Step	Description
Identify adjustment occasions.	The teacher decides when, during an instructional sequence, adjustment decisions should be made.
Select assessments.	The teacher chooses the formal or informal assessment procedures to be used for each adjustment occasion.
Establish adjustment triggers.	The teacher determines, in advance, what level of student performance will necessitate an instructional adjustment.
Make instructional adjustments.	The teacher makes any necessary adjustments.

Note. From Popham (2008, p. 53).

Figure 5. Formulating an adjustment trigger (Popham, 2008, p. 64).

In addition to homework (Vatterott, 2015) and quizzes, formative assessments come in many forms. Fennell, Kobett, and Wray (2015) identified five classroom-based formative assessments that teachers can use informally during instruction: (1) observations, (2) interviews, (3) show me (performance-based response), (4) hinge questions, and (5) exit tickets. Observations, interviews, and show me assessments involved the teacher observing their students working during classroom activity. Fennell et al. (2015) made the distinction, however, that observations were the most informal with the teacher not engaging in discourse with the students, interviews were more formal with the teacher specifically asking the students questions about their work, and show me assessments were the most formal with the teacher asking the students to demonstrate and explain their work while completing it. In addition to the different types of observations, Fennell et al. (2015) recommended the use of hinge questions and exit tickets as methods for in-class formative assessment. Hinge questions were those questions that measured important concepts which are essential for instruction to continue as planned (Fennell et al., 2015). Exit tickets, on the other hand, were informal, written responses to questions that students submit at the end of class which are used to measure the effectiveness of the lesson. No matter the level of formality, however, Fennell et al. (2015) argued that teachers need to adequately plan formative assessments to ensure effectiveness. For example, while observations were the most informal, Fennell et al. (2015) argued that teachers still need to plan for observations by identifying what they hope to observe and how they will identify it when they see it. While teachers should reserve exit tickets for one or two lessons per week, Fennell et al. (2015) suggested that teachers should use the other four types of classroom-based formative assessments at least once per class period.

To be truly formative, a teacher must use the feedback obtained from the assessment to modify and adapt instruction (Black & Wiliam, 1998; Chappuis et al., 2012; Tomlinson, 2014a; Wiliam & Thompson, 2008). Tomlinson (2014a) argued that

> Formative assessment is – or should be – the bridge or causeway between today's lesson and tomorrow's. Both its alignment with current goals and its immediacy in providing insight about student understanding are crucial to helping teacher and student see how to make near-term adjustments so the progression of learning can proceed as it should. (p. 11)

Teachers can modify instruction in several ways. For example, Popham (2009) offered two suggestions for modification based on the results of formative assessment. First, if the results of the assessment suggest that students have achieved an appropriate level of understanding, then the teacher might increase the pace of future instruction or remove plans for revisiting the content in future lessons. However, if the results suggest that students are struggling to obtain understanding, then the teacher should make necessary changes to their plans for future instruction.

Advocates of standards-based grading argued that effective formative assessments should be organized around well-defined learning targets and results should be reported to students based on those learning targets (Vatterott, 2015). Since formative assessments are meant to give students accurate information about their progress towards understanding those learning targets, students should be "expected to demonstrate the same level of skill or knowledge in the formative assessment that is expected in the summative assessment" (Vatterott, 2015, p. 63). Such formative assessments provide students with accurate feedback about their potential for success on any future summative assessment.

Differentiated Instruction

Advocates of standards-based grading recommended that teachers incorporate differentiated instruction as a response to the results of their continuous formative assessment (Vatterott, 2015). Bloom (1968) noted that students are not likely to master a concept if they require additional time to learn the concept, but do not receive that time. Therefore, Bloom argued that the teacher should tailor instruction to accommodate individual students instead of only considering whole classes or groups of students (see also Tomlinson, 2008). Tomlinson and Eidson (2003) defined differentiated instruction as “a systematic approach to planning curriculum and instruction for academically diverse learners” (p. 3). Two principles guide such instruction: (1) educators expect all students to achieve same, specific level of understanding of the content, and (2) student learning varies (Tomlinson, 2014b; Tomlinson & Eidson, 2003). Since the goal of standards-based grading is to provide all students an opportunity to achieve mastery of the content, advocates argued that differentiated instruction should be “embedded in the process of standards-based learning” (Vatterott, 2015, p. 59). That is, teachers should give students varied opportunities to learn as part of everyday instruction to account for the students’ differences in understanding.

Tomlinson and Eidson (2003) identified five classroom elements on which teachers could focus when differentiating instruction: (1) content, (2) process, (3) products, (4) affect, and (3) learning environment (see Table 7). In addition, they identified three student characteristics that teachers should consider when designing differentiated instruction: (1) readiness, (2) interest, and (3) learning profile (see Table 8). With respect to standards-based grading, differentiating with respect to content, process, and products based on readiness is essential (Vatterott, 2015). That is, teachers should focus on modifying how they teach and provide students opportunities to develop their own understanding based on the students’ current understanding of prerequisite knowledge as well as the knowledge the students still need to learn.

Table 7

Classroom elements to consider when differentiating instruction

Classroom Elements	Description
Content	What we teach and how we give students access to the information and ideas that matter.
Process	How students come to understand and "own" the knowledge, understanding, and skills essential to a topic.
Products	How a student demonstrates what he or she has come to know, understand, and be able to do as a result of a segment of study.
Affect	How students link thought and feeling in the classroom.
Learning Environment	The way the classroom feels and functions.

Note. Adapted from Tomlinson and Eidson (2003, p. 3).

Differentiating based on content means that a teacher should reflect on the essential mathematical concepts in the standards and curriculum to identify those concepts which are essential for each individual student (Tomlinson & Eidson, 2003). That is, based on the previous performance, the teacher should identify which concepts each student is struggling to understand and design instructional opportunities to address these concepts. For example, Tomlinson and Eidson (2003) recommended that a teacher provides students with additional learning opportunities, note-taking guides, or additional resources (e.g., instructional videos). Closely related to content, a teacher differentiates based on process by changing the way in which they ask individual students to think and reflect on the concepts they are learning (Tomlinson & Eidson, 2003). For example, a teacher might place students in homogeneous learning groups and design learning activities with different levels of difficulty based on the current level of understanding for each group which focus on the same concept. Finally, differentiating based on product means that the teacher offers students different opportunities to demonstrate their understanding (Tomlinson & Eidson, 2003). For example, Tomlinson and Eidson (2003) recommended that teachers provide students with alternative product formats that emphasize visual, auditory, and/or kinesthetic elements.

Table 8

Student characteristics to consider when differentiating instruction

Student Characteristic	Description
Readiness	The current knowledge, understanding, and skill level a student has related to a particular sequence of learning. It reflects what a student knows, understands, and can do today in like of what the teacher is planning to teach today.
Interest	What a student enjoys learning about, thinking about, and doing. Intended to help students connect with new information, understanding, and skills by revealing connections with things they already find appealing, intriguing, relevant, and worthwhile.
Learning Profile	A student's preferred mode of learning. The goal of learning profile differentiation is to help students learn in the ways they learn best – and to extend the ways in which they can learn effectively.

Note. Adapted from Tomlinson and Eidson (2003, p. 3).

To successfully differentiate based on readiness, effective ongoing formative assessments are essential (Guskey, 1997; Tomlinson, 2014b). Based on the results of these assessments, a teacher might offer alternative learning opportunities, called correctives, for students who are struggling to achieve mastery (Guskey, 1997). However, Guskey (1997) argued that, "to be successful, the correctives must be different from the original instruction" (p. 103). Therefore, opportunities for remediation and correction should offer students new opportunities for learning and growth. Once teachers provide the students with corrective instruction, students should have additional assessments opportunities to measurement their improved understanding (Gentile & Lalley, 2003).

Constructive Feedback

In addition to informing differentiated instruction, the formative assessments should provide both the teacher and the students with feedback about how the students are progressing in their understanding of the course content. Ramaprasad (1983) defined feedback to be "information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way" (p. 4). That is, feedback is information used to narrow the gap between a student's current understanding of a topic (i.e., the actual level) and the level of understanding which is expected by the teacher (i.e., the reference level). There are a couple important points about this definition that need highlighting. First, feedback cannot exist if information about the actual level, the reference level, or the gap is missing (Ramaprasad, 1983). Second, teachers much use such information to alter the gap in order to be considered feedback (Ramaprasad, 1983). In education, Black and Wiliam (1998) referred to these pieces of information as "recognition of the *desired goal*, evidence about *present position*, and some understanding of a *way to close the gap between the two*" (p. 143, emphasis in original).

Ilgen, Fisher, and Taylor (1979) characterized both goals and feedback as either specific or general (see Figure 6). Based on their classification, the feedback received will have varying degrees of interpretability and helpfulness. When educators define goals in general terms, specific feedback makes it difficult to measure performance and general feedback is difficult to interpret and apply. When goals are specific and feedback is general, teachers allow students to interpret that feedback based on their own frame of reference (Ilgen et al., 1979). In this situation, there is significant risk that students will misinterpret their performance as either being better or worse than reality (Ilgen et al., 1979). Ilgen et al. (1979) claimed that feedback is best implemented and understood with goals are well-defined and feedback is specific. As Tomlinson (2014a) noted,

> When feedback serves its instructional purpose, students are clear about the learning targets at which they are aiming, and they understand that assessments show how they are doing in reaching those targets. They trust that teachers will use the assessments to help them achieve, and they know that there will soon be follow-up opportunities for them to use the feedback in improving their performance. (p. 12)

		GOALS	
		SPECIFIC	GENERAL
FEEDBACK	SPECIFIC	Feedback is easily understood and applied to future performance.	Performance evaluation is difficult.
	GENERAL	Feedback is interpreted in terms of the performer's frame-of-reference.	Feedback is difficult to interpret and apply.

Figure 6. Interaction of goal and feedback specificity (Ilgen et al., 1979, p. 356).

Guskey (1997) argued that effective feedback is essential to successfully implementing standards-based grading and recommended that the feedback students receive from teachers needs to be regular, specific, diagnostic, and prescriptive. Specifically, Guskey (1997) claimed that feedback to students "should (1) reinforce precisely what was most important for them to learn in each unit of instruction, (2) recognize what students learned well, and (3) identify the specific concepts on which students need to spend more time" (p. 11). Such feedback comes from regular use of formative assessments (Guskey, 1997; Tomlinson, 2014a; Vatterott, 2105) and is non-punitive (Vatterott, 2015).

According to Hattie and Timperley (2007), "effective feedback answers three questions: where am I going?; how am I doing?; and where to next?" (p. 87). In order to answer these questions, they categorized four different types of feedback: (1) task, (2) process, (3) self-regulation, and (4) self. See Table 9 for definition of each type of feedback. Based on their research, task, process, and self-regulation feedback are the most effective at improving and motivating student learning. In contrast, self-level feedback "is rarely directed at addressing the three feedback questions and so is ineffective in enhancing learning" (Hattie & Timperley, 2007, p. 102).

Table 9

Categorization of types of feedback

Focus of Feedback	Description
Task	Focus is on how well the students performed or understood the task.
Process	Focus is on the main process needed to perform or understand the task.
Self-Regulation	Focus is on helping the students develop self-monitoring and regulation skills.
Self	Focus is on personal evaluations and affect about the learner.

Note. Adapted from Hattie and Timperley (2007).

Multiple Opportunities To Complete Summative Assessments

In contrast to a teacher using formative assessment to support instruction and student growth, advocates of standards-based grading argued that the purpose of summative assessments is to evaluate the students current understanding at the end of those learning opportunities (Vatterott, 2015). According to Burke (2010), "summative assessments serve as assessments *of* learning, because their purpose is to support the assignment of final grades or levels of proficiency related to course outcomes or state standards" (p. 23, emphasis in original). With respect to standards-based grading, the recommendation is that teachers organize summative assessments around learning targets where they report scores for each target (Vatterott, 2015). By recording results for each target, the teacher and students are better able to identify which standards the student still needs to master. Even after completing a summative assessment, advocates of standards-based grading argued that teachers should allow students to obtain remediation on those missed standards and retake the assessment to improve their results (Vatterott, 2015). Additionally, Gentile and Lalley (2003) recommended that the teacher should only record the students' best scores in the gradebook.

How Students Experience Learning: Intrinsic, Mastery, And Growth

Advocates of standards-based grading argued that when grades earned during the grading process are permanent parts of students' records, then it is no longer safe to make mistakes (Boaler, 2016; Vatterott, 2015). As a result, "students spend a great deal of energy avoiding imperfection and trying to look smart. This encourages deception, inhibits risk taking, and breeds a fear of failure and a false sense of shame" (Vatterott, 2015, p. 31). In contrast, standards-based grading, with its emphasis on non-punitive feedback and growth, fosters intrinsic motivation (Brookhart, 2011; McMillan, 2009), mastery goal orientations (McMillan, 2009), and growth mindsets (Boaler, 2016; Vatterott, 2015).

Incentive Theory

According to Ryan and Deci (2000), "to be motivated means *to be moved* to do something" (p. 54, emphasis in original). Individual's motivation can range from not motivated, or unmotivated, to very motivated. That is, people can have varying levels of motivation (Ryan & Deci, 2000). Ryan and Deci (2000) also claimed that people can have different orientations toward motivation. That is, different "underlying attitudes and goals that give rise to action" (Ryan & Deci, 2000, p. 54). Typically, we think of two different orientations: (1) intrinsic and (2) extrinsic. Ryan and Deci (2000) defined intrinsic motivation "as the doing of an activity for its inherent satisfaction rather than for some separable consequence" (p. 56). In contrast, Ryan and Deci (2000) defined extrinsic motivation as the doing of an activity "in order to attain some separable outcome" (p. 60).

Unlike intrinsic motivation, Ryan and Deci (2000) claimed that extrinsic motivation exists on a continuum based on the amount of internalization and integration an individual has experienced with respect to the specific activity (see Figure 7). Wentzel and Brophy (2014) defined internalization as "the transformations of an externally prescribed regulation or value into an internally adopted one" and integration as "the process through which internalized regulations and values become integrated into the self" (p. 74). With respect to the classroom, through the process of internalization, students assimilate externally endorsed values from their teacher into personally or internally endorsed values. Once internalized, the students then further integrate those values into their personal beliefs and motives. Specifically, Ryan and Deci (2000) argued that "internalization and integration are the processes through which extrinsically motivated behaviors become more self-determined" (p. 65). That is, they become more intrinsically motivated. Figure 7 shows the stages of the continuum of internalization from external regulation to integration and eventually intrinsic motivation.

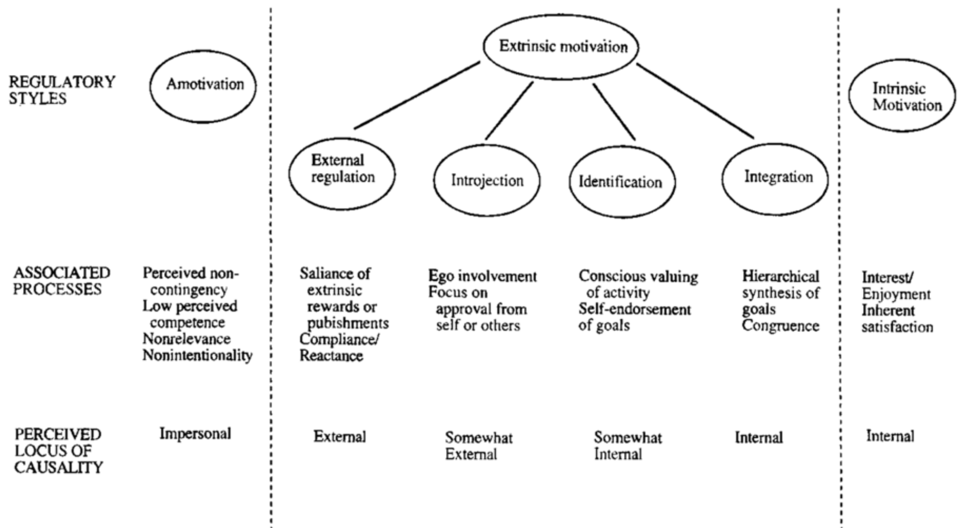

Figure 7. A taxonomy of human motivation (Ryan & Deci, 2000, p. 61).

Rewards can have a significant impact on the process of internalization and integration of regulation and values (Ryan & Deci, 2000). Such rewards can be classified as verbal and tangible (e.g., prizes, physical objects, privilege) (Deci, Koestner, & Ryan, 2001; Ryan, Mims, & Koestner, 1983). Table 10 provides a classification of rewards. Deci et al. (2001) found that, except for some verbal rewards, all types of rewards have a significant negative impact on student intrinsic motivation. According to Deci et al. (2001), the implementation of rewards has two important aspects: informational and controlling. Specifically, "the *informational* aspect conveys self-determined competence" and "the *controlling* aspect prompts an external locus of causality" (Deci et al., 2001, p. 3, emphasis in original). Since tangible rewards offer little information about student performance and competence, they have low informational value. In addition, teachers typically use tangible rewards as a classroom management tool and, therefore, these types of rewards have high controlling value to both teacher and student. Thus, they have a significant negative impact on intrinsic motivation (Deci et al., 2001; Ryan et al., 1983). Advocates of standards-based grading, argued that grades are examples of performance-contingent tangible rewards and reducing their use will cause students to become more intrinsically motivated to learn (Gentile & Lalley, 2003; Vatterott, 2015).

Table 10

Classification of verbal and tangible rewards

Reward Type	Reward Type	Description/Expectation
Verbal	Verbal/Written	Verbal or written feedback about performance.
	Task-Non-Contingent	Given to people for being present for an activity, but does not require engaging in the activity.
	Engagement-Contingent	Given to people for engaging in the activity, but does not require completing it.
Tangible	Completion-Contingent	Given to people for completing the activity, but does not require a particular level of performance.
	Performance-Contingent	Given to people for completing the activity to a specific level of excellence or criterion.

Note. Adapted from Deci et al. (2001) and Ryan, Mims, and Koestner (1983).

Recall that Hattie and Timperley (2007) argued that self-level feedback offers students little information about their performance. Therefore, verbal rewards at the self-level typically have low informational value and students often perceive them as having high controlling value (Deci et al., 2001). Thus, such rewards also have a significant negative impact on intrinsic motivation. Deci et al. (2001) argued that only verbal rewards which “contain explicit positive performance feedback” (p. 3) have the potential to increase intrinsic motivation.

Achievement Goal Theory

According to Meece, Anderman, and Anderman (2006), student behavior is purposeful and directed toward obtaining certain goals related to developing and demonstrating competence. Ames and Archer (1988) identified two types of goal orientations: performance and mastery. When a student has a performance goal orientation, they become concerned with others judging them as being capable of achieving the goal (Ames & Archer, 1988, p. 260). Specifically, students become focused "on demonstrating high ability relative to others, striving to be better than others, and using social comparison standards to make judgments of ability and performance" (Meece et al., 2006, p. 490). In contrast, when a student has a mastery goal orientation, they become focused "on developing one's abilities, mastering a new skill, trying to accomplish something challenging, and trying to understand learning materials" (Meece et al., 2006, p. 490).

Elliot (1999) further divided both performance and mastery goal orientations into two directions: approach and avoidance. An approach motivation is one driven by a "positive or desirable event" and an *avoidance* motivation is one driven by a "negative or undesirable event" (Elliot, 1999, p. 170, emphasis in original). Consequently, Elliot and McGregor (2001) posited four types of goal orientations: performance-approach, performance-avoidance, mastery-approach, and mastery-avoidance. For example, students with a performance-approach goal orientation would seek favorable judgements of their ability, while students with a performance-avoidance goal orientation would avoid negative judgements of their ability (Meece et al., 2006).

Pulfrey, Buchs, and Butera (2011) showed that when students anticipated graded feedback, they increased adoption of performance-avoidance goal orientations. Such results are concerning because other research suggested that students with performance goal orientations, regardless of direction, were more likely to engage in cheating behaviors (Anderman & Midgley, 2004; Murdock & Anderman, 2006). In contrast, McMillian (2009) argued that receiving non-punitive feedback, as recommended by advocates of standards-based grading, increases the likelihood of students adopting a mastery-approach goal orientation (McMillian, 2009). Ames and Archer (1988) showed that when students perceived their class as fostering mastery-approach goals, "they were more likely to report using effective learning strategies, prefer tasks that offer challenge, like their class more, and believe that effort and success covary" (p. 264).

Mindset Theory

In addition to fostering intrinsic motivation and mastery-goal orientations, advocates of standards-based grading claimed that such a philosophy can have a significant impact on students' mindsets towards learning (Vatterott, 2015). Dweck (2006) defined an individual's mindset as their personal view of their intellectual ability (Dweck, 2006). Mindset theorists argued that an individual's mindset can have a significant impact on behavior and motivation (Dweck, 2006; Yeager & Dweck, 2012). While given different names, the literature focused on two types of mindsets: entity and incremental (Yeager & Dweck, 2012). See Table 11 for descriptions of each mindset with respect to academic settings. Students with an entity theory of intelligence, called a fixed mindset, view intellectual ability as an object of which people have an unchangeable amount (Dweck, 2006; Yeager & Dweck, 2012). In contrast, students with an incremental theory of intelligence, called a growth mindset, view intellectual ability as an object that people can grow and develop with time and effort (Dweck, 2006; Yeager & Dweck, 2012).

Table 11

Academic mindsets for those with an entity versus incremental theory of intelligence

	Entity Theory	Incremental Theory
Goals	Look smart	Learn
Value of effort, help, and strategies?	Signal that they lack natural talent	Essential to success and growth
Response to challenge.	Tendency to give up	Work harder and smarter
Changes in grades during times of adversity.	Decrease or remain low	Increase

Note. Adapted from Yeager and Dweck (2012, p. 303).

According to Dweck (2006), students with a fixed mindset seek opportunities to prove themselves as smart and avoid those opportunities where they perceive that they are deficient in intelligence. As a result, Boaler (2016) cautioned that traditional grading systems cause students to fear making mistakes in mathematics and, as a result, fosters a fixed mindset in mathematics. Such results are concerning because students with a fixed mindset tend to study less for exams, more often consider cheating, become less willing to work with others (Dweck, 2006), and struggle to accurately interpret performance feedback (Mangels, Butterfield, Lamb, Good, & Dweck, 2006).

Boaler (2016) and Vatterott (2015) argued that standards-based grading practices change the perspective of learning and intelligence from being “fixed and permanent” to being a process that is “growing with time.” That is, standards-based grading fosters a growth mindset (Boaler, 2016; Vatterott, 2015). Mangels et al. (2006) found that, in general, students with growth mindsets experienced significantly greater gains in knowledge than their fixed mindset counterparts. In classrooms that implemented growth mindset messages, the ability gap between high- and low-ability students narrowed (Dweck, 2006).

Implications For The Current Research Study

As noted in the previous section, standards-based grading has the potential to have significant impact on student motivation, goals, and mindset (Boaler, 2016; Brookhart, 2011; Gentile & Lalley, 2003; McMillan, 2009; Vatterott, 2015). However, to accomplish this potential, teachers must change their grading and instructional practices (Boaler, 2016; Brookhart, 2011; Gentile & Lalley, 2003; McMillan, 2009; Vatterott, 2015). Specifically, teachers need to be able to (1) address a large number of academic standards using flexible timing and high-cognitive demand tasks (Brookhart, 2012; Gentile & Lally, 2003), (2) implement non-punitive, continuous formative assessment which offers informative, performance-focused feedback (Guskey, 1997; Popham, 2008; Tomlinson, 2014a), (3) use formative assessment feedback and data to inform differentiated instruction which addresses the continuing learning needs of all students (Guskey, 1997; Tomlinson, 2014a), and (4) offer multiple opportunities to demonstrate improved understanding (Gentile & Lalley, 2003; Vatterott, 2015).

Boesen et al. (2014) found that reform efforts in mathematics are often met with resistance. Specifically, instead of modifying their own teaching philosophies to adjust for new ideas, teachers tended to assimilate the new ideas into their existing teaching philosophies (Boesen et al., 2014). The result of such assimilation was that teachers claim to abide by the ideals of the new system while still maintaining the tenets of the old system. Researchers documented such results with respect to the implementation of standards-based grading (Brodersen & Randel, 2017; Gentile & Lally, 2003; Welsh & D'Agostino, 2009).

As noted in the previous chapter, there exists a gap in the current literature with regards to implementing standards-based grading practices in middle school mathematics classrooms. Much of the current research has focused on quantitative measures of alignment with standardized assessment scores (Bradbury-Bailey, 2011; Deddeh et al., 2010; Hochbein & Pollio, 2016) and surveys of teachers' grading practices (Hochbein & Pollio, 2016; Tierney et al., 2011). There exists a need for qualitative research focused on classroom implementation of standards-based grading practices and its impact on classroom instruction (Brodersen & Randel, 2017; Pollio & Hochbein, 2015). This study sought to fill this gap in the literature by exploring the impact of standards-based grading systems on the practices of middle school mathematics teachers.

CHAPTER III

METHODOLOGY

The purpose of this qualitative, multicase study was to describe middle school mathematics teachers' teaching practices as they implemented stands-based grading. Specifically, the study answered the following research questions:

Q1 How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?

Q1a What is the nature of the mathematical tasks selected by middle school mathematics teachers for instruction?

Q1b How do middle school mathematics teachers facilitate mathematical instruction?

Q1c What is the nature of classroom discourse as facilitated by middle school mathematics teachers during instruction?

Q1d How do middle school mathematics teachers utilize assessment strategies?

Q2 How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?

Q3 What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?

Q3a How do middle school mathematics teachers' own teaching philosophies impact their implementation of standards-based grading practices?

Q3b How do other stakeholders (e.g., school district personal, parents, students) impact middle school mathematics teachers' implementation of standards-based grading practices?

Guided by an interpretive, theoretical perspective, the researcher implemented a multicase study design (Stake, 2006) focused on four teachers' implementation of standards-based grading practices within their middle school mathematics classrooms.

Data collection consisted of a combination of interviews and classroom observations. The data were analyzed at the both the case and cross-case levels for themes related to the purpose of this study. Peer examination and member checking where among the strategies used to ensure increased trustworthiness and rigor.

Theoretical Perspective And Epistemology

According to Crotty (1998), a theoretical perspective is “the philosophical stance informing the methodology and thus providing a context for the process and grounding its logic and criteria” (p. 3). The study was guided by an interpretive, or a constructivist, theoretical perspective (Creswell, 2013; Merriam, 2009). According to Merriam (2009), interpretive research “assumes that reality is socially constructed, that is, there is no single, observable reality. Rather, there are multiple realities, or interpretations, of a single event. Researchers do not ‘find’ knowledge, they construct it” (pp. 8-9). From an interpretive perspective, researchers acknowledge that participants “develop subjective meanings of their experiences” (Creswell, 2013, p. 24). As a result, it is important for researchers to “look for the complexity of views” and “to rely as much as possible on the participants’ views of the situation” (Creswell, 2013, pp. 24-25).

An important component of a theoretical perspective is its epistemological stance; that is, “the theory of knowledge embedded in the theoretical perspective” (Crotty, 1998, p. 3). As part of taking an interpretive perspective, the researcher accepted the constructionist view that knowledge and meaning are constructed rather than acquired (Crotty, 1998). That is, as humans engage with their world, they interpret their surroundings and through this interpretation they construct meaning. As a result, truth can be neither objective nor subjective, but rather "one's way of making sense of the world is as valid and worthy of respect as any other" (Crotty, 1998, p. 58).

The use of an interpretive theoretical perspective and the corresponding epistemological stance of constructionism are consistent with a theoretical lens of cultural-historical activity theory because the researcher focuses on understanding the "specific contexts in which people live and work in order to understand the historical and cultural settings of the participants" (Creswell, 2013 p. 25). Furthermore, from the lens of cultural-historical activity theory, learning is the result of individually perceived contradictions within an activity system (Engeström & Sannino, 2010). Through the resolution of such contradictions, individuals modify and expand their personal knowledge and meaning of their activity system.

Multicase Study Design

The purpose of this qualitative research study was to describe middle school mathematics teachers' teaching practices as they implemented standards-based grading. According to Merriam (2009), the purpose of qualitative research is "to achieve an *understanding* of how people make sense out of their lives, delineate the process (rather than the outcome or product) of meaning-making, and describe how people interpret what they experience" (p. 14, emphasis in original). To achieve the goal of this research study, the researcher implemented a multicase study research design (Stake, 2006). Researchers utilize a multicase study design when they seek to understand a phenomenon through the exploration of representative cases (Stake, 2006). With respect to this research study, multicase study design allowed for the development of a detailed description of middle school mathematics teachers' teaching practices as they implemented standards-based grading through the analysis of individual case records of the phenomenon.

The Quintain

According to Stake (2006), multicase study research focuses on understanding "an object or phenomenon or condition to be studied" (p. 6) which he calls the *quintain*. To understand the quintain, researchers need to study multiple cases which represent manifestations of the quintain. Specifically, the goal is to "study what is similar and different about the cases in order to understand the quintain better" (Stake, 2006, p. 6). In this multicase study, the quintain of focus was middle school mathematics teachers' teaching practices as they implemented standards-based grading practices. Specifically, the quintain consisted of the teachers' uses (or non-uses) of assessment, feedback, grading practices, and instructional design as well as the challenges or support they encountered while utilizing these grading practices. With respect to cultural-historical activity theory, this consisted of the interactions between the subject's (i.e., the teacher's) standards-based grading rules and (1) other rules (e.g., classroom norms), (2) the community (e.g., students, parents), (3) the division of labor (e.g., student engagement, expectations of feedback), (4) instrumentation (e.g., the lesson plan, use of assessment), and objects (e.g., lesson objectives).

The Cases

While the quintain is the phenomenon under investigation, researchers examine it through the study of individual cases which are examples or manifestations of the quintain (Stake, 2006). That is, "the individual cases should be studied to learn about their self-centering, complexity, and situational uniqueness" (Stake, 2006, p. 6).

According to Stake (2006), the individual cases should be the initial focus of attention similar to individual case studies. Only after the individual cases are understood in depth, can the research focus change to understanding the quintain (Stake, 2006).

According to Creswell (2008), "a case study is an in-depth exploration of a bounded system (e.g., an activity, event, process, or individuals)" (p. 476). Creswell further defined *bounded* to mean that the case can be "separated out for research in terms of time, place, or some physical boundaries" (p. 476). Stake (1995) classified a case as "an integrated system" (p. 2). Since the purpose of this study was to illuminate a particular issue (the quintain), the research identified instrumental cases for in-depth study (Creswell, 2008; Stake, 1995). That is, the researcher sought cases that helped to understand teachers' implementations of standards-based grading as part of middle school mathematics instruction.

Research Sample

The researcher invited four middle school mathematics teachers who self-reported using standards-based grading practices in their classroom to participate in the multicase study (Morgan & Powers, 2018). The boundary for each case consisted of each teacher's instruction pertaining to a single class over the course of five class periods. In order to protect the anonymity of participants, the researcher gave each participant a pseudonym. In addition, the author used the gender-neutral terms they, them, and theirs when describing and discussing the teachers' practices.

Stake (2006) identified three criteria for selecting cases for a multicase study: "(1) is the case relevant to the quintain?; (2) do the cases provide diversity across contexts?; and (3) do the cases provide good opportunities to learn about complexity and context?"

(p. 23). Using data from Morgan and Powers (2018), the researcher identified four middle school mathematics teachers to invite as participants: Mx. Taylor Brown, Mx. Reilly Johnson, Mx. Alex Williams, and Mx. Jamie Miller. All four teachers self-reported that their school districts mandated that they use standards-based grading as part of their instruction of middle school mathematics. That is, their cases were relevant to the quintain.

With respect to diversity, complexity, and context, the teachers provided several similarities and differences which contributed to the ability to uncover variation within the quintain. Each teacher worked in a middle school (i.e., Grades 6 through 8) within three different large, suburban school districts located on the front-range of Colorado; with Mxs. Brown and Miller working for the same school district, but at different middle schools. While each teaching middle school mathematics, the teachers differed with respect to their overall teaching experience, experience with standards-based grading, and student grade-level.

Data Collection

When conducting research from an interpretive theoretical perspective, Creswell (2013) recommends that researchers utilize interviews with open-ended questioning to understand the participant's constructed reality. As the participant shares their perspectives, the researcher makes "an interpretation of what they find, an interpretation shaped by their own experiences and background" (Creswell, 2013, p. 25). The interpretation is the researcher's perspective on "the meanings others have about the world" (Creswell, 2013, p. 25).

To address the research questions, the researcher obtained institutional review board (see Appendix A) and school district approval to collect several forms of data as part of this multicase research study. For each teacher participant, data consisted of (1) an initial, semi-structured interview, (2) lesson summaries and reflections, (3) lesson documentation, (4) classroom observations, and (5) a final, semi-structured interview. See Table 12 for a summary of each data source. In addition to classroom observations, the use of interviews and self-recorded summaries and reflections allowed for the documentation of the participants' interpretations of their constructed reality (Creswell, 2013).

Semi-Structured, Initial Interview

The first data source was a semi-structured interview. The purpose of the initial interview was to gather information about each teacher's teaching practice in general as well as with respect to their implementation of standards-based grading. In addition, the interview included questions to better understand the classroom norms, students, and typical lesson design for the class under consideration. See Appendix B for the interview protocol. See Table 13 for an alignment between the research questions and a subset of the tentative interview questions. The semi-structured, initial interview was approximately 60 to 90 minutes in length and was audio-recorded. Mxs. Brown's and Johnsons' interviews took place at the teachers' schools during the school day prior to the beginning of classroom observations. In Mxs. Williams' and Miller's case, a face-to-face interview was not possible due to scheduling constraints. As a result, the interviews took place over the phone prior to the beginning of classroom observations.

Table 12

Summary of data sources

Data Source *Type*	Focus *Research Questions Addressed*	Duration *Timing*
Semi-Structured, Initial Interview *Audio Recorded*	Understanding the teacher's teaching practice, implementation of standards-based grading, classroom norms, and up-coming lesson plans. *Q1, Q1a, Q1b, Q1d, Q2, Q3, Q3a, Q3b*	60-90 minutes *Prior to the start of week of classroom observations.*
Lesson Plan Summary *Audio Recorded*	Understanding the teacher's plans for the up-coming lesson including goals, lesson structure, and assessments. *Q1, Q1a, Q1b, Q1d, Q2, Q3, Q3a, Q3b*	5-10 minutes each *Prior to the start of each observed lesson.*
Lesson Reflection *Audio Recorded*	Understanding the teacher's perceived success (or failure) of the implemented lesson and the potential for the lesson to inform up-coming lessons. *Q1, Q1a, Q1b, Q1d, Q2, Q3, Q3a, Q3b*	5-10 minutes each *Following each observed lesson.*
Lesson Documentation *Photocopied*	Understanding the resources the teacher uses when planning for and implementing instruction. *Q1, Q1a, Q1b, Q1d*	*Submitted to the researcher with appropriate Lesson Plan Summary and Lesson Reflection.*
Classroom Observations *Video Recorded*	Capture classroom instruction with the primary focus being on the teacher's interactions during instruction. *Q1, Q1a, Q1b, Q1c, Q1d*	60-90 minutes each *5 consecutive class periods for a single class*
Semi-Structured Final Interview *Audio Recorded*	Understanding the teacher's teaching practice, implementation of standards-based grading, and perceived success of previous lessons. *Q1, Q1a, Q1b, Q1d, Q2, Q3, Q3a, Q3b*	90-120 minutes *Following the final classroom observation.*

Table 13

Alignment between the research questions and tentative initial interview questions

Overarching Research Question	Tentative Initial Interview Questions
How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?	What would a typical class period in your classroom look like? How do you plan for instruction? What resources do you use?
How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?	What polices and/or norms, if any, do you have related to this practice? How do you think standards-based grading has impacted your classroom instruction?
What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?	What is your perception of support and/or resistance in implementing this practice by administrators? Other teachers? Parents? Students?

Lesson Summaries And Reflections

The second data source was a summary and reflection of each lesson taught by the teachers. For each of the observed lessons, each teacher audio-recorded a lesson plan summary prior to teaching the lesson and a lesson plan reflection following the lesson. During each lesson plan summary, the teachers discussed their goals and plan for upcoming lesson, their rationale for the flow of the mathematical concepts, and other lesson plan decisions. See Appendix C for the guiding summary and reflection questions. During each lesson plan reflection, the teachers discussed their perception of the success of the lesson, their use of any assessments, potential mid-instructional decisions, and potential impacts the lesson will have on future lessons. The teachers' summaries and reflections ranged from three to eight minutes in length. See Table 14 for an alignment between the research questions and a subset of the summary and reflection guiding questions. The researcher gave each teacher an audio-recorder to use at a time and place when recording their summaries and reflections is possible.

Table 14

Alignment between the research questions and tentative summary and reflection questions

Overarching Research Question	Summary and Reflection Guiding Questions
How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?	*Guided Summary Questions:* What are you planning to do in the upcoming lesson? What are your mathematical learning goals for the upcoming lesson? *Guided Reflection Questions:* How did the lesson you just taught go as compared to what you had planned to do?
How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?	*Guided Summary Questions:* What role (if any) will your use of standards-based grading play in the upcoming lesson? *Guided Reflection Questions:* What role (if any) did your use of standards-based grading play in the previous lesson?
What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?	*Guided Summary Questions:* What contributed to your decision to present the upcoming lesson in this way?

Lesson Documentation

The third data source entailed lesson documentation. For each of the observed lessons, the researcher collected photocopies of all relevant lesson artifacts including, but not limited to, the lesson plan, worksheets, handouts, lesson assignments, and lesson assessments. The documentation served as a secondary resource to better understand the intended lesson design for the planned lesson. The researcher collected the lesson documents at the same time as the lesson summaries and reflections between lessons. Since students were not research participants, data collection did not include any student work other than work that students presented to the whole-class during regular classroom instruction.

Classroom Observations

Classroom observations were the fourth data sources used in this study. To document classroom instruction, five consecutive lessons for almost every teacher were audio- and video-recorded as well as observed by the researcher. In Mx. Williams' case, only four classroom observations were possible due to scheduling conflicts with school-related testing and assemblies. For each observation, the researcher set-up and ran a video-recording device in an unobtrusive position in the classroom. In addition, the teacher wore a lapel microphone to capture their classroom interactions. Each class period ranged from 60 to 90 minutes.

The purpose of the recorded lessons was to capture classroom instruction with the primary focus being on the teacher's interactions during instruction. The video-recording device was set up in the back of the classroom with the goal of recording the teacher while minimizing the capture of identifiable information for the classroom students. In addition to audio- and video-recording classroom instruction, the researcher recorded written field notes of the observations. The purpose of the notes was be to gather additional information about classroom instruction.

Semi-Structured, Final Interview

The fifth data source was a final interview whose purpose was to gather additional information about the observed lessons and the teacher's implementation of standards-based grading. See Appendix D for the interview protocol. See Table 15 for an alignment between the research questions and a subset of the tentative interview questions. The semi-structured, final interview were 90 to 120 minutes in length and were audio-recorded. Each interview took place at the teacher's school during the school day following the final classroom observation.

Table 15

Alignment between the research questions and tentative final interview questions

Overarching Research Question	Tentative Final Interview Questions
How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?	Did your lessons go as planned? If you were to teach these lessons again, what would you do differently?
	What types of assessments to you give? What are their purposes?
	What feedback do you give students about their understanding of content?
	How do you work with students who need remediation?
	You are using standards-based grading, how do you determine and/or define which standards to address?
How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?	How do you assign student grades? What type of evidence do you include in grade determination?
What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?	You once used a different grading practice, how has standards-based grading changed your teaching practice?

Data Analysis

In multicase study research, Stake (2006) recommended conducting two stages of data analysis: (1) case study analysis and (2) cross-case analysis.

Individual Case Study Analysis

When conducting a case study, Patton (2002) recommended first developing a case record and then reducing the data into a case study report.

Developing the case record. According to Patton (2002), “the case record includes all the major information that will be used in doing the final case analysis and writing the case study” (p. 449). With respect to this study, the case records for each teacher included (a) coded transcriptions of each interview, (b) coded transcriptions of each lesson summary and lesson reflection, (c) partitioned lesson observations, (d) lesson documentation, and (e) detailed descriptions of the case.

Interviews, summaries, and reflections. The researcher used a transcription software, Transana, to transcribe each interview, lesson summary, and lesson reflection. Using the cultural-historical activity theory framework (see Table 1), the researcher coded the transcriptions using a qualitative coding software, NVivo. By using this framework, the researcher identified contradictions and tensions within the teacher’s perceived activity system (Engeström, 2015). That is, the researcher identified aspects of the teacher’s practice that informed instructional decisions at the macro- and micro-levels of teaching (Wells, 1996). After coding, the researcher developed thick descriptions of the teacher’s practice based on each interview, lesson summary, and lesson reflection.

Lesson observations. Unlike the interviews, summaries, and reflections, the researcher did not transcribe each lesson observation. Instead, the researcher used a qualitative data analysis software, NVivo, to analyze the video-recordings. As part of that analysis, given the length and scope of the lessons, the researcher partitioned each lesson into smaller, more manageable episodes and sequences (Wells, 1996). Wells (1996) argued that “classroom events are best understood as ‘actions’ which, organized sequences of activities and tasks, enact the Practice of Education” (p. 76). To accomplish this goal, Wells (1996) recommends dissecting the discourse into nested components (see Figure 8). See Table 16 for descriptions of each component.

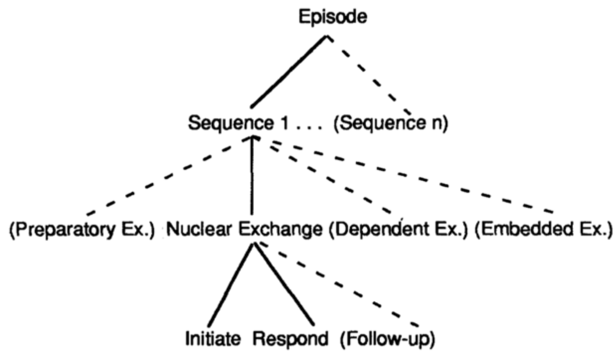

Figure 8. Organization of spoken discourse (Wells, 1996, p. 78).

According to Doyle and Carter (1984), academic tasks "provide a central classroom structure that governs student information processing. A description of such tasks should provide, therefore, insight into how the curriculum is realized on a daily basis in classrooms" (p. 131). The researcher partitioned the classroom discourse into episodes based on enacted mathematical tasks. Recall that Stein et al. (1996) defined mathematical tasks "as a classroom activity, the purpose of which is to focus students' attention on a particular mathematical idea" (p. 460). Specifically, a mathematical task "is not classified as a different or new task unless the underlying mathematical idea toward which the activity is oriented changes" (Stein et al., 1996, p. 460). Once partitioned into episodes, the researcher further divided each episode into sequences (Wells, 1996).

Table 16

Descriptions of the components of spoken discourse

Component	Mathematical Classroom
Episode	The talk that occurs in the performance of an activity or one of its constituent tasks.
Sequence	Includes a single nuclear exchange and any exchanges that are bound to it.
Exchange	Reciprocally-related moves which constitutes the minimal unit of spoken discourse. Consists of an *initiating* move, a *response* move, and, in some cases, a *follow-up* move.
Nuclear Exchange	Can stand alone, independently contributing new content to the discourse.
Bound Exchanges	Not free-standing, but depend on the nuclear exchange in some way.
Dependent Exchange	Some aspect of the nuclear exchange is developed through further specification, exemplification, justification, and so on.
Move	The smallest building block. Consists of one instance of an individual's spoken language. Does not constitute discourse by itself.

Note. *Adapted from Wells (1996).*

According to Wells (1996), the level of sequence is the most useful unit of analysis when analyzing discourse and joint activity. As a result, the researcher focused on the sequence level when coding the classroom observations. Specifically, the researcher coded each sequence of discourse for (a) instruction type (i.e., whole-class, small-group, and individual instruction), (b) teacher moves, (c) opportunities for feedback, and (d) task cognitive demand. To assist in coding, the researcher used resources from the literature including the mathematical cognitive demand framework (Table 4 and Table 5), a teacher moves framework (Figure 3), and a feedback framework

(Table 9). After coding, the researcher developed thick descriptions of the classroom observations.

Developing the case study report. Following coding and analysis of each data component, the researcher combined the thick descriptions from each data source into a single case study report for each participant (see Chapters IV through VII). Since the purpose of the individual cases is instrumental, the focus of the individual case studies was "to help us understand the phenomena or relationships within it" (Stake, 1995). With such cases, Stake (1995) recommended foregoing "attention to the complexity of the case to concentrate on relationships identified in [the] research questions" (p. 77). Therefore, the researcher developed synthesized, individual case study reports that focused on better understanding their contribution to better understanding the quintain (Stake, 2006) by using the cultural-historical activity theory framework as a guide (see Table 1).

Cross-Case Analysis

According to Stake (2006), "given the binding concept – a theme, issue, phenomenon, or functional relationship that strings the cases together – the [researcher has] an obligation to provide interpretation across the cases" (p. 39). Guided by the research questions, the researcher combined each individual case study report, with specific examples from the data, to highlight similarities and differences across the cases (Patton, 2002; Stake, 2006). See Chapter VIII for a report of the cross-case analysis. The goal of the cross-case analysis was to develop a thick description of the quintain (Merriam, 2009; Stake, 2006). Therefore, the researcher focused on the similarities and differences (a) among the individual teacher's enactment of standards-based grading and learning, and (b) among the teachers' enactments of standards-based grading and the recommendations from the literature.

Trustworthiness And Rigor

Merriam (2009) identified credibility, consistency, and transferability as key characteristics of trustworthiness and rigor in qualitative research. She defined credibility, or internal validity, as the extent to which the observed data represents reality. In contrast, she defined consistency, or reliability, as the extent to which the "findings of a study are consistent with the data presented" (Merriam, 2009, p. 222). Finally, transferability was defined as the extent to which a report provides enough information so that the reader can decide whether to apply the findings to another situation (Merriam, 2009). This research study used peer examination, member checking, triangulation, an audit trail, bracketing, and thick descriptions as methods for improving the credibility, consistency, and transferability of the proposed study.

Peer Examination

Merriam (2009) recommended using peer examination as a method for ensuring credibility of the research findings. Peer examination refers to the process of asking peers "to examine the data and to comment on the plausibility of the emerging findings" (Merriam, 1995, p. 55). Merriam (2009) suggested that this is a naturally occurring relationship between graduate student and research advisor. Based on this recommendation, the researcher worked with their research advisor to discuss "the congruency of emerging findings with the raw data and tentative interpretations" (Merriam, 2009, p. 229). Specifically, the two met regularly during both the data collection and data analysis phases to discuss issues and emerging themes.

Member Checking

In addition to peer examination, Merriam (2009) recommended member checks as a strategy for ensuring credibility. Member checks refer to the process of soliciting feedback from participants on emerging findings to determine their level of agreement with themes and conclusions. From an interpretative perspective, member checks are an essential component of qualitative research to make sure that emerging themes represent reality from the perspective of the participant (Creswell, 2013). By allowing the participants to review their case report, the researcher was better able to ensure that the case report accurately reflected the experiences of each participant.

To accomplish member checks, the researcher sent each participant their individual case reports to review for accuracy in representation. Of the four teachers, only Mx. Brown responded to the request for feedback; noting that the case report did "a very faithful job of representing what happens in [their] classroom." They offered several clarifications related to the grade-level content of their course (i.e., they taught Grade 7 and 8 content as opposed to only Grade 8 content) and the phrasing of classroom discourse (i.e., using numerator and dominator which discussing simplification, and changing the word "shapes" to "objects" when discussing a simulation activity). Mx. Brown also suggested using the specific name for a classroom management technique; however, the researcher and advisor decided against making this change as it did not add understanding of the case and increased the risk to anonymity of the participant.

Triangulation

In addition to seeking outside input as strategies for ensuring credibility, a research design utilizes data triangulation as an additional method for ensuring credibility

(Merriam, 2009). Specifically, a research design utilizes observations, document analysis, and multiple interviews as data collection methods. When results or emerging themes are supported from each data source, then the themes have been triangulated. That is, the themes have increased credibility (Merriam, 2009). The researcher used all data collected when developing emergent themes and triangulated the findings across data sources.

Audit Trail

To ensure consistency, Merriam (2006) recommended keeping an audit trail; that is, "a detailed account of the methods, procedures, and decision points in carrying out the study" (p. 229). Schwandt (2007) claimed that an audit trail serves two purposes:

> It can be used by the [researcher] as a means of managing record keeping and encouraging reflexivity about procedures, and…it can be used by a third-party examiner to attest to the use of dependable procedures and the generation of confirmable findings on the part of the inquirer. (p. 13)

As part of completing this study, the researcher kept a detailed audit trail over the course of the study. Based on the recommendation of the literature (Merriam, 2009; Schwandt, 2007), the researcher kept a journal and organized electronic folder documenting the data generated during the study and notes about the process of conducting the study including documentation of important decisions made during data collection and analysis.

Bracketing

Creswell (2013) cautioned that the researcher's personal background shapes their interpretation of their participants' realities. Therefore, it is essential that interpretive researchers bracket their personal biases and experiences to reduce their impact on final conclusions (Creswell, 2013; Merriam, 2009). That is, researchers should reflect on and document "their biases, dispositions, and assumptions regarding the research to be undertaken" (Merriam, 2009, p. 219). By documenting personal biases, the researcher provides the reader with information which helps them understand "how the individual researcher might have arrived at the particular interpretation of the data" (Merriam, 2009, p. 219). Based on this recommendation, the researcher recorded their initial and final perspectives and beliefs about the implementation of standards-based grading practices in middle school mathematics.

Initial perspective. As a former middle school mathematics teacher, I acknowledge that issues exist with the use of traditional, letter-grade grading systems. It was my experience that there was little coordination between teachers who taught the same course and a lack of consistency between the courses taught by the same teacher. In the courses I taught, my grading policies typically changed from year to year as I attempted to find a structure that worked *best* for my practice, my school, and my students. I am both intrigued by and skeptical of the ideas and promises made by the literature pertaining to standards-based grading (e.g., Boaler, 2016; Gentile & Lalley, 2003; Heflebower, 2014; Vatterott, 2015). It is plausible to me that, when implemented as recommended by the literature, standards-based grading practices can have a significant impact on students' motivation and achievement. However, I question whether it is possible to implement all the recommendations as well as cover the seemingly large amount of content required in mathematics classes.

Additional bracketing. After completing this research study, I continue to question whether it is possible to implement all the recommendations of the literature. Over the course of data collection, data analysis, and writing, I regularly discussed my developing conclusions about these practices with my advisor. We focused on maintaining a non-deficit view (Spangler, 2014) of the four teachers who participated in my study. That is, we focused on ensuring that the data spoke for itself as opposed to allowing our personal biases and beliefs about teaching influence how the teachers' practices were presented. This included removing the desire to offer recommendation and criticism of the teachers' practices. While I may have acted differently under similar circumstances, I believe that the teachers in this study acted in the best way they could, under the constraints they were given, to ensure the best possible learning opportunities for their students.

Rich Description

Finally, Merriam (2009) recommended including sufficiently rich descriptions of cases as a method of improving transferability. It is the view of the researcher that this final report includes "enough description to contextualize the study such that readers will be able to determine the extent to which their situations match the research context, and, hence, whether findings can be transferred" (Merriam, 2009, p. 229).

Structure Of The Results

The results of this research study were separated into five chapters. The first four chapters focus on the individual case reports for each teacher. The order of the chapters are based on the order of data collection with Mx. Brown first, followed by Mx. Johnson, Mx. Williams, and finally Mx. Miller. This order highlights for the reader the order in which the researcher gathered the data used to develop their conclusions about standards-based grading. Finally, the fifth chapter presents the cross-case analysis. The report of the results is followed by a chapter which includes a discussion of the results grounded in the literature with recommendations for practice and future study.

CHAPTER IV

MX. TAYLOR BROWN

One of the things I really like about standards-based grading is it makes me focus on my mental concept of the student's mastery and it eliminates concern about things like points.
~ Mx. Taylor Brown ~

Prior to the start of class, Mx. Taylor Brown is at the door greeting students as they enter the classroom. As the students find their seats, they start on the daily warm-up problems which are projected on the board. The students sit in multiple rows configured in pairs, called shoulder partners. As the students start the warm-up, Mx. Brown prepares for the upcoming lesson.

Mx. Brown's case highlights how an experienced teacher who is new to a school district implements the district's standards-based grading policies, with little professional development, by accommodating the policies within their personal teaching and assessment strategies, and the constraints of a middle school mathematics classroom.

Subject

Mx. Brown has been teaching for over five years but transitioned into their current position the previous year. Having originally pursued a career in another field, Taylor first taught at the high school level and, with this new position, taught middle school during data collection. Over this transition, they believe they "evolved as a teacher." In that sense, Mx. Brown finds "that you really have to strike a balance between having boundaries, but also taking care to develop the relationships with the kids." Taylor values preparing the students "for the practical aspects of [the students'] future education."

Mx. Brown's view of teaching is strikingly different from their past views of teaching. As a student, they went through school with the "idea that the teacher should know everything; the teacher should make every effort to deliver all [the knowledge] into [them] bucket fashion." That is, the teacher should have given Taylor "as much knowledge as possible." In contrast, Mx. Brown now views teaching students as "not the filling of a bucket, but the lighting [of] a flame." In this sense, they believe that, to learn, "you need to be able to figure things out for yourself."

In their previous teaching positions, Mx. Brown used a traditional system of grading. However, after transitioning to their current position, the school district mandated that all teachers, including Taylor, use a standards-based grading system.

Community

Mx. Brown expects their students to be independent, "self-reflective" learners who take responsibility for learning the course material. In addition to their students, other teachers as well as the greater school district community impact Taylor's implementation of standards-based grading in their middle school.

Students

The students in the observed accelerated seventh grade class are "pretty high-power students." As an "accelerated" course, the Mx. Brown teaches the students seventh and eighth grade content. Preparing the students to take Algebra the following school year is part of Taylor's goals for the course. Taylor hypothesizes that "math would be the most commonly hated subject and the most challenging subject" for middle schoolers. They try to “show the kids how what they're learning will be useful and [they've] taken what [they] would consider to be an honest approach with that.”

Being an accelerated course, the pace is faster, and the number of concepts covered during the year is significantly greater than the students encountered in previous years. Mx. Brown notes that the class follows a “more rigorous pace [and] a lot more rigorous content.” As a result, many of the students in the class, approximately two-thirds, “have threes; whereas, [the students] are used to having fours.” Taylor believes that it is their job “to teach the kids at a grade level proficiency and when [the students are] there, then theoretically [Taylor] can be happy that [they're] meeting the goal.” As a result, Mx. Brown perceives their students as being frustrated by not achieving high scores; whereas, Taylor perceives the students as meeting the grade level goals and is satisfied with the students’ progress in the course.

When working during class, Mx. Brown expects students to take ownership in their learning. Taylor's goal is to “coach” students to be “equal partners in the learning [where] the teacher is not solely responsible.” Mx. Brown also expects their students to be “self-reflective.” For example, before obtaining help from Taylor, students should “have the problem already written down [and] check google to try to see what it means.”

Mx. Brown asks students to work effectively individually as well as in pairs or small-groups. Taylor assigns students their seats “for management” purposes. In addition to their daily assigned seats, the students occasionally work in pre-assigned partner pairings. During activities where students work in the pairings, the students rotate to new pairings every 6 minutes. Each student had 12 different pre-assigned pairings that stayed consistent throughout the year. At the beginning of the school year, the students picked their partners. Taylor claims this process was “part of the empowering [of their students] to make wise decisions.” The students in the observed class could be “a high energy and chatty bunch.” When students misbehave, the students “might have to come visit [Mx. Brown] at lunch, might owe [Mx. Brown] a little bit of time after class, or might owe extra homework.”

Other Teachers

According to Mx. Brown, there was “very limited team planning in the math department, [but] not because [the teachers] don't get along.” Mx. Brown identifies different lunch periods and different preps as limiting factors on the teachers' abilities to co-plan. With respect to the observed course, Mx. Brown co-planned the initial sequence of the course with one other teacher. However, in Taylor’s view the collaboration would be ineffective as the other teacher taught a different population of students. Specifically, the other teacher taught sixth grade students who had not previously learned traditional sixth grade mathematics; whereas, Mx. Brown’s students had completed a traditional sixth grade mathematics course. Due to this additional course, Taylor believes their students have more prerequisite knowledge. As a result, Taylor perceives the other common course teacher as “taking the approach of a little bit slower and deeper.” In contrast, Taylor plans “to stick with the plan that [the two teachers had come] up with” at the beginning of the school year.

The School District

Mx. Brown claims that they received “no formal training” on standards-based grading practices; preparation merely consisted of seeing district-level descriptions for the district’s 4-point rubric. In addition, Taylor uses the district-issued gradebook to document and report students' grades. While Mx. Brown is aware of past tension with respect to standards-based grading and its development in the school district, those issues occurred before they started working in the district. Taylor believes that there were "a lot of people [who were] generally dissatisfied with certain things about [standards-based grading]." For example, Mx. Brown noted that some frustration exists with respect to the interpretation of grades by students and parents. In response to this frustration, Taylor thinks it is "next to impossible to get kids and parents to stop equating A as four, B as three, C as two, D as one, and F as zero." Mx. Brown also recalls discussions about switching the gradebook system to use a "power function" where the gradebook would "weight the things towards the end of the term more heavily." However, Taylor believes that "whoever was in charge of the gradebook in the district couldn't quite get that to work."

Mx. Brown refers to the middle school as existing on an "isolated island" within the school district because the elementary and high schools are not doing standards-based grading. In their view, this lack of consistency as students progress through the grades is "a pretty strong argument" for not doing standards-based grading at the middle school level. Taylor thinks that "the kids don't like it." Adding that "it confuses the heck out of the parents," but in the parent's defense conceded, "how much time do [parents] have to put into understanding the system?" Taylor believes that "consistency is important and the lack of it kind of drives [Taylor] a little bit nuts." However, Mx. Brown does not value one system over the other. In particular, if the high school were to adopt a standards-based grading system, then they would be okay keeping the system. Though, they "don't see that really happening."

Instruments

Based on their understanding of the school district's standards-based grading policies, Mx. Brown created their own process, or "mental calibration," for supporting instruction and assessing students' understanding of course content. As part of their implementation of the district-mandated standards-based grading policies, Mx. Brown utilizes a district-designed gradebook, a standards-aligned curriculum, and multiple types of assessments.

Standards-Based Grading Policies

Mx. Brown claims they did not receive "much formal training on the way the [standards-based grading] system is meant to be implemented." They partially attribute this to "low turnover in the building" because, with only a few new teachers, the perception might have been that training was not necessary. Instead, they have "done [their] best to kind of look at some things online and look at the descriptors and do what [they] feel is best." While "somewhere along the line [they] kind of saw the descriptors for the different numbers" in the district-mandated rubric, Mx. Brown grades by using their "mental calibration for assigning the scores" based on their recollection of those descriptors.

Mx. Brown believes that "the idea of standards-based grading is to show what a student knows." They like standards-based grading because it puts the focus on "mastery" of concepts as opposed to obtaining points. That is, "it's not about trying to get a worksheet [turned] in [at the] last minute to get 15 points and bump up to the next letter grade." Because of this shift in focus, Taylor claims that when they observe students during class, they no longer think about letter grades, but rather they think about how they would score them on a specific concept. Mx. Brown views the scores they gave their students as “very reflective of what the students know.”

According to the district-mandated gradebook, students’ overall grades are divided into three categories. Summative assessments account for 70% of the student's grade and formative assessments account for 29% of the student's grade. The remaining 1% of a student’s grade is intended to be the district-mandated “work habits,” measured across four work habits, including: (a) I'm productive, (b) I participate, (c) I exhibit a positive mental attitude, and (d) I'm prepared. Mx. Brown admits that they did not “often enter those grades.” However, they note that “it can be nice for the parents to get a read on some of their kid’s 'more soft' skills [like] attitude, preparedness, participation, that sort of thing.” Ideally, with more time, Taylor suggests that they could see themself entering a work habit grade once a month for approximately nine total grades per year.

When deciding on a grade for non-computer-based work, Mx. Brown follows a specific process. First, Taylor aligns the assessment to one of the five common core standards or “gradebook buckets.” Then, since each standard has “a whole lot of specific math skills” connected to it, Taylor asks themself if the student's work is “a three for [those] particular skills?” They claim to “not really think about the nebulous, giant bucket of an overall standard.” That is, they think about specific skills related to the standard as opposed to thinking about the overarching standard when assigning grades. As the year progresses, Mx. Brown reinterprets the meaning of a score of three relative to the concepts covered in class as well as what concepts coming up next. That is, a score of three means that a student is “proficient with those particular [current] math skills [and are] on track to do whatever is coming next”. A three during first semester, from

Taylor's perspective, means that the student is ready for the next unit; whereas, a three during second semester means that the student is ready for the next course (i.e., Algebra).

Mx. Brown has a "mental" concept for the meaning of each score on the 4-point grading scale when grading non-computer-based work. A four means that a student can teach the concept to a student. A three means that a student can "do most of the problems with confidence." That is, there might "be some that [the student] doesn't know how to do, like the harder ones, but [the student] can do most of the problems with confidence." A student might earn a two if the student can "definitely do some of the problems, but [the student] has obvious stumbling points." A score of one means that a student "could maybe do the easiest ones, but all in all [the student] feels confused."

Because of their interpretation of score of one, Mx. Brown views any student work as evidence of some degree of student understanding. That is, if a student shows any work, then Taylor claims they have evidence that the student's grade is at least at a one. As a result, it is Mx. Brown's practice to not typically give students a score of zero. Taylor only gives a zero if they do not "have enough evidence to make an educated decision" about the students understanding. Only students who are "chronically truant" earn a zero because those students likely did not take the assessment and did "not bother to come in and make it up." Taylor believes that their practice of not giving zero scores differs from other teachers in the middle school. For example, they claim that some teachers give students a zero if the student answered most of the items incorrectly. In contrast, Taylor gives a student a one, if the student did "all the problems and [got] every single one wrong." In this case, Taylor argues that the student showed "evidence that [the student had] limited proficiency [and they] don't know the math."

Mx. Brown's scoring practices change when assigning scores for online, computer-based assessments. The need for this change stems from the fact that the computer system scores students' work based on a percentage of correct items. When looking at individual students' online scores, Taylor claims the system does not support identifying specific concepts students have mastered and which the students are struggling to understand. As a result, Mx. Brown scores students' online performance based on a percentage-to-score conversion. Specifically, Taylor considers a four to be 90% and above, 70% and above is a three, 50% and above is a two, and anything less that 50% is a one. If a student is within 5% of the next score, then Mx. Brown gives the student a "point five." For example, if a student scores an 85%, then Taylor enters a 3.5 in the gradebook for the online assessments.

Mx. Brown replaces grades when a student improves their performance on an assessment. Mx. Brown believes that test retakes work well for students who are motivated; however, they claim that very few students retake assessments. They recall one teacher suggesting that 10% to 20% of their students would retake an assessment to which Mx. Brown note that they "wish [they were] getting 10 to 20% for retakes." Taylor hypothesizes that, because of the lack of consequences for failing a class, grades in middle school did not matter for most students.

When recording retakes, Mx. Brown replaces a lower score with the higher score. The decision to not keep old scores in the gradebook is because Taylor believes it is not fair to a student who is earning a four to be "stuck with the ones and twos." For example, if a student scores a one on a quiz, but then scores a three on a retake for that quiz, then Taylor replaces the student's grade with a three. While Mx. Brown acknowledges that a student might lose knowledge over time, they argue that “the last time [they] saw [that student, the student] had a level three proficiency.”

Mx. Brown has mixed feelings about providing opportunities for students to retake the assessments. On one hand, for students who “tank a test or just don’t do as well,” they can work on correcting their understanding and retake the assessment. Taylor claims that this helps to alleviate students' feelings of test anxiety. On the other hand, Mx. Brown notes that retakes are a “pain in the butt, because [they] have to write two versions of every test instead of one.”

District-Issued Gradebook

The school district designed the online gradebook that Mx. Brown uses to document and report student grades. As a result, the grade distribution and alignment options were already programmed into the online platform. When entering grades into the gradebook, Taylor only has to “select whether an assignment is a work habit, formative, or summative” and select the standard to which the assignment is aligned.

The structure of the online gradebook consists of overarching domains and substandards. In the previous year, Mx. Brown was required to select one of the substandards with which to align each assignment. When aligning at the substandard level, Taylor admits that there were times in which an assignment might cover more than one skill, but they would only align it with one substandard because they did not want it to double count. At the start of the current school year, however, the district decided to change the gradebook so that teachers only had to align at the standard level. Taylor explains that the “only decision [they have to make] is does this fit in with ratios and proportions, expressions and equations, number system, geometry, or probability and stat." Mx. Brown admits that they "honestly don't put a lot of thought into it because it's like these giant buckets and it's very easy to tell which bucket an assignment belongs in."

When adding assignments to the gradebook, district personnel directed Taylor to use a work around to correct for a weighting issue in the gradebook. Suppose that Mx. Brown wants an end-of-chapter test to be weighted more heavily than other items in the gradebook. They were "told not to put [the test] in as eight points, [but rather to] make two assignments each worth four points." For example, the gradebook includes two assignments called "Chapter 13 Test" for which students received the same score out of four points and one assignment called "Section 13.1 Quiz." In this case, the test is contributing twice as many points to the student's overall grade than the quiz. Taylor claims that this process "effectively doubles the weight at least within the [aligned] standard" in the gradebook. However, it does not double the weight with respect to other standards in the gradebook.

In Mx. Brown's view, grades are "supposed to be reflective of what the kid knows [and] are supposed to be a reporting tool for [them], for the kids, for the parents." Taylor believes that the system is getting there. They admit that there are "some mathematical inadequacies" which are "really tough to get around." However, Mx. Brown thinks "it all kind of works out in the end anyway." Taylor does not use their grade book "as a way to tell which students are deficient in which things;" however, they claim that the grading system is satisfying their needs and they are "seeing in the gradebook what [they] need to see."

Curriculum Resources

Mx. Brown uses the district-issued curriculum that includes a textbook and online learning platform. The curriculum is Taylor's “roadmap because it is common core aligned.” Mx. Brown plans to complete 12 to 13 chapters over the course of the school year. The chapter sequence was “designed to prepare [the students] for the Algebra class [the students] will go into next year [and] secondarily to get the content that is the focus of the PARCC test.” Mx. Brown admits that they trust that the common core standards and the curriculum “align themselves pretty well.” Taylor also uses “the PARCC test as a target for the level of rigor [their students] should be able to answer.”

In addition to the printed textbook resources, Mx. Brown utilizes the curriculum’s online platform for in-class activities, assessments, and online homework assignments. When looking at the online homework, students can see the questions they answered correctly and incorrectly. Taylor hypothesizes that some students take advantage of the multiple-choice structure by continually clicking answers until they answer the question correctly. The online platform includes videos to support students in learning the material. Mx. Brown recommends that students watch those videos as an alternative to waiting for Taylor to come help.

Assessments

Mx. Brown utilizes summative assessments as well as graded and ungraded formative assessments. While the gradebook structure distributed grades so that summative assessments counted for 70% and formative assessments counted for 29%, Taylor practice of not assessing student work habits changed the distribution to “basically

70% and 30%" respectively. As a part of daily classwork, some formative assessments do not count towards a student's grade.

Summative assessments. Mx. Brown categorizes quizzes and tests as summative assessments. On occasion, Taylor writes their own tests, while on other occasions they use the online assessments. Mx. Brown estimates about two-thirds of the students prefer the teacher-written tests because the students perceive the online tests as more difficult.

Each assessment receives a different weight based on its significance in measuring student understanding. A mid-chapter quiz goes in the gradebook as one summative grade, and end-of-chapter tests go in as two summative grades. That is, end-of-chapter tests contribute twice as much towards a students' grade as a mid-chapter quiz. An end-of-semester final counts as three summative grades. If Mx. Brown thinks the test is "a really, really, really central end-of-chapter test," then it might count as three summative grades. Taylor defines "central" as something being important for success in Algebra (e.g., solving equations or graphing linear equations). However, Mx. Brown considers a concept in geometry to be "not as central because [the students] are hardly going to touch it in Algebra."

Formative assessments. Mx. Brown administers both graded and ungraded formative assessments. Homework serves as a graded formative assessment in the sense that (a) Taylor records students' performance on the homework in the gradebook as a formative assessment score using the same conversion for grading computer-based scores outlined above, and (b) Taylor can obtain information about class' overall understanding of specific questions and concepts. The online homework allows Mx. Brown to see individual student's performance on each question, but it requires Taylor to wait for several pages to load. Instead, Taylor focuses their attention on the overall class reports. Mx. Brown argues that they are "teaching the class, [they're] not teaching an individual student, so if [they're] teaching the class and [the] itemized report" shows that the class is struggling on specific numbers, then that is where they are going to spend their time.

In general, most of the problems on homework assignments consist of the current content, but Mx. Brown often includes some review. For example, a homework assignment focused on properties of exponents might also include a problem focused on practicing the order of operations. During classroom observations, however, Mx. Brown did not assign the students any homework as a reward for good behavior. They note that this decision was "sort of as a reward and sort of because right now [they] are really busy." By not assigning homework, Taylor is able to make the students "feel good while also just saving [themself] a little bit of time." In the future, Mx. Brown plans to assign shorter homework assignments focused on practicing the properties and processes that the class is learning at the time.

In addition to online homework, Mx. Brown utilizes two in-class activities as ungraded, informal formative assessments. First, the students complete daily warm-up problems at the start of the class. Mx. Brown typically assigns a few questions related to the current topic and one or two questions "to bring back old concepts." The purpose of including previously taught material is to "keep it fresh with kind of the dual purpose of helping [the students] remember it for Algebra and helping them remember it for the PARCC testing." The warm up serves as a "quick check" for Mx. Brown to determine where the students are "in comparison to each other."

End of class, rapid-fire activities serve as the second type of ungraded, informal formative assessment Mx. Brown uses. These activities serve as a quick check-in for Taylor and consist of “very quick questions about the concept [the class] has been doing.” During these activities, Mx. Brown calls on students in a random order to answer questions. They hypothesize that the class will get through 15 to 20 questions in 3 to 5 minutes. The purpose of the activity is to “act as a quick formative check [and to] act as a bunch of extra repetitions.”

Remediation And Feedback

When preparing for a retake, Mx. Brown recommends that students watch YouTube videos and practice problems out of the book. Taylor is also available to give students “extra support at lunch or after school.” Mx. Brown expects students to work with them when it comes to obtaining help. It is Taylor's “philosophy” that they could not “force” help on the students. They are there “to help [students] as much as [those students want] to be helped, but [Taylor] isn’t going to do it for [those students].”

In general, Mx. Brown finds it logistically difficult to organize students for remediation during class because all the seats in the classroom are full. However, on occasion, they design classroom activities to support remediation. For example, to review for assessments, Taylor creates stations around the room for students to complete tasks geared towards reviewing specific concepts. However, during these activities, Mx. Brown does not spend time telling students which topics to review because they are “depending on” each individual student to determine the topics that student needs to work on. Therefore, the activity provides the opportunity for remediation on specific topics, but the students have the choice to opt out of participating on remediation for that topic in favor of a different topic.

Object

A typical class period in Mx. Brown's class starts with the students completing a set of warm-up problems individually, followed by discussing the agenda for the day, then engaging in some collection of activities with a focus on the current topic, and concluding with a rapid-fire review activity. To maintain the students' attention, Mx. Brown segments the middle portion of the lesson into two or three parts. During this portion of class, Taylor asks students to work individually, in small-groups, or as a whole-class on activities that have students working either on worksheets or on online activities.

The observed lessons focus on developing properties of exponents. Mx. Brown teaches this topic via "an inquiry-oriented approach" by starting with an activity designed to develop the properties. Using "an inquiry style activity" to start the unit is "a little bit atypical" for Mx. Brown who views teaching the lessons in this way as " a learning experience.". After the entry activity, Taylor attempts to see how much of the knowledge their students can develop without the teacher's influence.

Summary Of First Observation

As students enter the classroom for the first observation, the agenda for the day directs the students to prepare for a warm-up problem and to get out an activity that the class started the previous class period. After a quick welcome, Mx. Brown gives students time to work on the four warm-up problems. See Figure 9 for a recreation of these problems. When finished, students turn in the warm-up problem sheets on Taylor's desk.

1) Simplify: $x + x + x + x + x =$

2) True or False: $5x = x^5$

3) Simplify: $x \cdot x \cdot x \cdot x \cdot x =$

4) $x^{12} \div x^3 =$

Figure 9. Warm-up problems from the first observation of Mx. Brown's class.

Following the students' individual worktime on the warm-up tasks, Mx. Brown briefly thanks the students for good behavior with a recent substitute teacher, and announces that, as a reward, the students will not have any assigned homework for the week. After this announcement, Taylor leads the class in a brief discussion of the solutions to the warm-up problems. This discussion consists of asking students the solution of each problem and addressing any student confusion. In response to student frustration with their incorrect answers, Mx. Brown reminds the class that "some math concepts you learn faster, some you'll learn slower and that will not be at the same rate as your friend learns it and that is totally fine."

Before moving onto the activity started the previous class period, Mx. Brown discusses the results of a recent online test. By show of hands, most of the class indicate that they want to do a test retake. Mx. Brown notes that "makes [their] life simple" and that they will offer an in-class retake during the following week. In response to a student question about the grading scale, Taylor reminds the students that "a 90 and up is a four, 70 and up is a three, 50 and up is a two, [and] if you're within 5% of the next range you get point five. So, an 85 is a 3.5, 65 is a 2.5, 45 is a 1.5." The class average is a 72, and, since it was a challenging test, from Taylor's perspective, this means that the students are "proficient as a class." In particular, the students "are accomplishing what [the] government wants [the students] to accomplish… and [the students] are getting ready for the PARCC test." Mx. Brown tells the students that the retake will be a different version of the online assessment, but will have similar questions.

As the class transitions into the "exponents investigation worksheet" from the previous lesson, Mx. Brown tells the students to be thinking about how the students could show and explain their reasoning. Taylor references work from a warm-up problem as an example of "showing" reasoning. See Figure 10 for a recreation of the work to "show" the simplification of x^{12}/x^{3}. The work shows that x^{12} can be written as the repeated multiplication of 12 *x*'s, and x^{3} can be expressed as the repeated multiplication of three *x*'s. Then, division of an *x* by itself results in a "giant one" which is represented by the rectangles. The work shows three rectangles to represent three *x*'s in the numerator cancelling out with three *x*'s in the denominator completing the division problem for the simplified answer x^{9}.

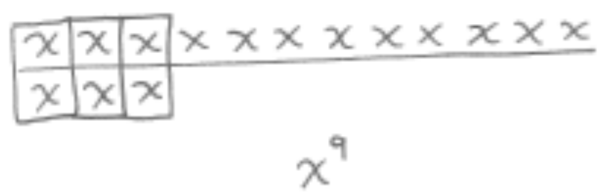

Figure 10. Example work demonstrating how students "show" their reasoning.

For the next 45 minutes, the students work in pairs and small-groups on completing an exponents investigation worksheet which focuses on guiding students towards developing properties of exponents. These properties include, among others, properties related to division of exponential expressions with the same base and properties related to an exponent of zero. As students work, Mx. Brown monitors the students and offers guiding hints and suggestions when students express confusion.

During this time, Taylor tells the students that "the purpose of this [activity] is not for [the teacher] to simply tell [the students] the answers in a notes-style fashion." Instead, "the purpose is for [the teacher] to give [the students] little leading hints to see what [the students] can come up with on their own." After time to work, Taylor brings the whole class together to present the solutions to the activity. To motivate the discussion, Mx. Brown announces that the class is "about ready to practice these skills, but before [the students] can practice [the students] have to know what [they're] doing." As Taylor presents the solutions, they tell students that the chapter is "not about memorizing rules," such as $a^m/a^n = a^{m-n}$. Instead, if the students forget the rule, then the students should simply expand the expression and cancel to get the simplified answer.

Following the presentation of solutions to the worksheet, Mx. Brown asks the students to get out mini whiteboards at their desks to participate in a "whiteboard practice" activity. Taylor tells the students to use the exponents exploration worksheets to help on the practice, but the goal is to eventually not need the worksheets as a guide. During this activity, Mx. Brown projects, one at a time, practice problems on the overhead. The students then write and show the solution on an individual whiteboard. Taylor looks around the room and announces the solutions they see. Afterwards, Mx. Brown then reveals the correct solution on the overhead. During the activity, the class simplifies eight different exponent expressions. See Figure 11 for a recreation of a couple of these expressions. To end the class period, the students create "bingo boards" for an exponent bingo game Mx. Brown plans for the following class period. Students fill out a grid with exponents of the form x^n where n ranges from -13 to 10. Taylor encourages the students to practice the exponent rules that evening even though the students do not have formal homework.

$$(x^2)^5 = \qquad \frac{w^8}{w^2} = \qquad j^0 = \qquad a^5 \cdot a^3 \cdot a^{10} =$$

Figure 11. Sample of expressions used for the whiteboard practice activity during the first observation of Mx. Brown's class.

Summary Of Remaining Lessons

As the lessons progress throughout the week, instruction focuses on providing opportunities for students to practice and develop intuition about the properties of exponents. The lessons include individual, small-group, and whole class activities in online and paper environments.

Second observed lesson. Before students start the warm-up, Mx. Brown checks in with the students to determine how the students feel about their understanding of the properties of exponents via thumbs up or down. There is a mixture of student responses, and Taylor tells the students they plan to slow the chapter down. The goal of slowing down the chapter is to see if Mx. Brown can "move more of [the students] from threes to fours." Following this brief discussion, the students complete the warm-up problems which consist of two problems from the current chapter and two problems that "came from long ago, November or so." See Figure 12 for a recreation of these problems. After students turn in the warm-up tasks, Mx. Brown asks students to present solutions to the warm-up problems on the board. As the class discusses the solutions to the warm-up problems, Taylor encourages the students to discuss amongst themselves any disagreements about the solutions.

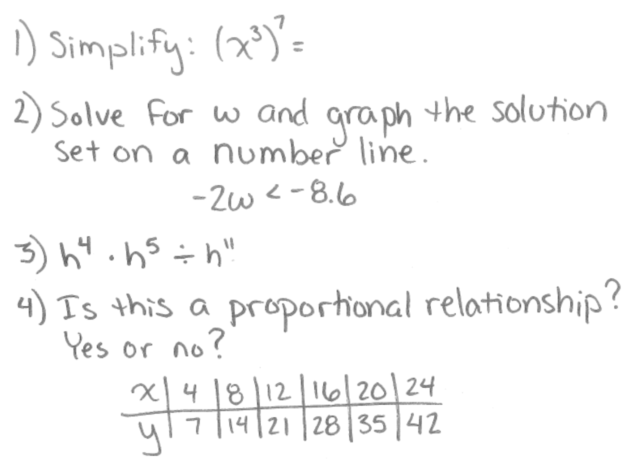

Figure 12. Warm-up problems from the second observation of Mx. Brown's class.

Following a quick discussion of the agenda for the day, the class transitions to an order of operations review worksheet which asks students to simplify expressions with multiple operations. See Figure 13 for examples of these expressions. During this activity, Mx. Brown asks the students to work for 5 minutes individually, and then the students can work with a shoulder partner. After giving the pairs time to work on the review, Taylor presents the solutions to the worksheet at the board. Mx. Brown shows their work and provides explanation for any problems about which the students express confusion. Following this discussion, the class plays two rounds of exponents bingo during which Mx. Brown projects an expression for the students to simplify and find the equivalent expression on their premade bingo boards. Each round ended after about five students correctly achieve a "bingo." To verify a "bingo," the students announced which answers contributed to their five-in-a-row, and Taylor confirmed that the expressions the student said corresponded to simplifications of the expressions they presented during play.

1. $2 + 1 \cdot 4^2 - 12 \div 3$
2. $8^2 \div 16 \cdot 2 - 5$
3. $7(9 - 2) + 6^2 \cdot 10 - 8$

Figure 13. Sample of expressions from the order of operations review completed during the second observation of Mx. Brown's class.

After a quick formative check-in with the students to gauge their comfort with the properties of exponents, Taylor ends the class with a rapid-fire activity with a focus on properties of exponents and combining like terms. The expressions used in both the bingo and rapid-fire activities were similar to those used in the whiteboard practice activity from the day before.

Third observed lesson. Due to perceived high energy coming from the students, Mx. Brown starts the third day with "a few moments of calm" during which the students sat in quiet reflection. Following student work time on the daily warm-up problems, Taylor tells students to log into the online learning platform to work on problems from the sections focused on properties of exponents. As students work in the online environment, Mx. Brown invites volunteers to write the solutions to the warm-up problems on the board. After the solutions are written on the board, Taylor asks the students to take a break from the online practice to participate in a discussion of the warm-up problems. During the discussion, Mx. Brown asks the students to discuss and explain errors and corrections amongst the students. After this discussion, Taylor directs the students to return to the online exponents practice by working individually. During this practice, Mx. Brown encourages the students to choose problems that the students do not already know how to do. That is, the students should "be doing those problems that [the students] perceive to be moderately challenging." The problems should "guide learning" and those problems might be different for each student. After 20 minutes of practice, the class transitions to an online activity with a focus on developing additional understanding of the properties of exponents. The class concludes with a rapid-fire activity with a focus on the properties of exponents and an exit ticket reflection on effective learning strategies.

Fourth observed lesson. As directed by the agenda projected on the board, the students acquire a textbook as they walk into the classroom. The class does not complete a set of warm-up problems on this day. Instead, Mx. Brown leads a discussion of important strategies for working together including being open to the ideas of others. Following this discussion, Taylor tells the students that they will be continuing to practice the properties of exponents. On this day, the class works on practice problems out of a textbook. The students first work for 6 minutes individually and then rotate to a new partner pair every 6 minutes. As the students work, Mx. Brown monitors the students' progress and offers help and guidance as needed. The class period concludes with a computer-driven simulation of real-world objects ranging in size from subatomic particles to celestial bodies. The purpose of this activity is to motivate the real-world application of properties of exponents and scientific notation.

Fifth observed lesson. To start the class period, the students complete the daily warm-up. After the students turn in the warm-up, Mx. Brown directs students to grab the textbooks and asks for student volunteers to write solutions to the warm-up problems on the board. During the discussion of the solutions, Mx. Brown tells the class that a student with a "level four understanding" might forget the rule, but should be able to "use their tool to recreate the rule." After this discussion, Taylor announces that, before moving onto new content (i.e., scientific notation), they want to allow the students more time to practice properties with negative exponents. Mx. Brown perceives the students as struggling with problems that include negative exponents. As a result, the plan for the day is to work, similarly to the previous day, on practice problems from the book that focus on negative exponents. The students first work for 6 minutes individually and then rotate to a new partner pairs every 6 minutes. As the students work, Mx. Brown monitors the students' progress and offers help and guidance as needed. Due to student misbehavior, Taylor decides not to end the class with a game of exponents bingo. Instead, the students work individually on problems of their choice in the online platform. At the end of class, Mx. Brown announces that, going forward, the class will be moving onto new content. Therefore, if the students perceive that they are still confused on the properties of exponents, then they will need to work on practice problems outside of class or use other resources (e.g., YouTube) to develop their understanding.

Tensions

Mx. Brown does not specifically cite any issues with standards-based grading that directly impacts their implementation. However, in discussions of their practice, Taylor recalls struggles with limited professional development, concerns about converting students grades to letter grades, and issues related to time management.

Limited Professional Development

Mx. Brown claims they never received formal training focused on how to implement standards-based grading. While they recall seeing a rubric or recommended descriptions of student grades, they grade using a "mental calibration" of what they believe the 4-point grading scale should measure with respect to student understanding. Taylor notes that they had to do “research” on their own time to find support for implementation, but that they only have so much time to spend researching practices. In the end, Mx. Brown relies on the support of other teachers through informal, passing conversations. For example, to correct for weighting errors in the online gradebook, other teachers suggested the solution to add assessments multiple times to increase the assessments’ overall weight in the gradebook. However, Taylor admits that these teachers also have different perspectives on how to implement standards-based grading practices.

Grades And Gradebook

Mx. Brown expresses concern about converting student grades into letter grades. Taylor questions whether there was “any equation or translation that [would] be totally accurate or is there always going to be [a student] who is given too much benefit of the doubt or undercut somewhere.” Mx. Brown does not teach a high school credit course and is thankful that they did not have to worry too much about converting student grades into letter grades. Though, they note that the process was “a little bit worrisome.” In particular, they are concerned that the students in the district will struggle to be competitive for scholarships when measured against their peers who are not being measure used standards-based grading practices.

Concerns related to grade conversions partially stem from Mx. Brown’s acknowledgement of “flaws” in the gradebook system. For example, one assignment in geometry would be equivalent to 20 assignments in expressions and equations because the gradebook averages scores within each standard and then averages the standards equivalently to determine an overall score. As a result, the student's grade might be misleading because the one geometry assignment is impacting the student's score more than a single assignment in expressions and equations. However, when Taylor reflects on their students' grades, they believe that the grades are reflective of what the students know.

Time And Expectations

When it comes to grading and instructional time, Mx. Brown regularly makes choices about how to spend their time. When scoring students' work habits, Taylor admits to not entering those grades because their time is better spent on other priorities. When teaching the standards, Mx. Brown makes decisions about what content to emphasize. In Taylor's view, they are "basically getting them ready for success in the Algebra class." As a result, Mx. Brown believes their students will "be able to learn the geometry just fine no matter how much or little [Taylor] touches on that now." Mx. Brown uses information about homework from the whole class rather than from individual students to decide how to spend instructional time during class.

Summary

Mx. Brown's case highlights how an experienced teacher, with limited professional development, modifies their teaching practice to accommodate their interpretation of a school district's grading practices while also modifying those grading practices to fit within their already established beliefs about teaching and learning. Taylor's goal of preparing their students for success in Algebra and their understanding of the curriculum resources drives their implementation of standards-based grading. This perspective directly influences how Taylor determines and assesses students' understanding of “grade level” content, and how they allocate their time for instruction and assessment by determining which topics to emphasize and how much time to spend on those topics.

CHAPTER V

MX. REILLY JOHNSON

As a math teacher, I'm like, this is the most amazing thing ever because I can look at ratios and proportions, expressions, equations, geometry, and I can be like, this student is really good in all of these, but man, the geometry is killing them.
~ Mx. Reilly Johnson ~

As students enter the classroom, Mx. Reilly Johnson is at the door greeting and welcoming the students to class. The directions telling students to engage in a conversation about their homework assignments is projected on the board at the front of the classroom. The students sit in groups of four around the classroom. Each group consists of one student assigned to one of the following roles: group facilitator, task manager, recorder/reporter, and resource manager. Reilly has the lesson goal, essential question, and plan for the day written on a side board for students to follow. As students engage in conversation, Mx. Johnson starts the daily homework check.

Mx. Johnson's case highlights how a teacher incorporated district-developed resources into their implementation of instruction and standards-based grading practices. This case demonstrates how a teacher will interpret and modify district and school policies and practices to support their beliefs about teaching and learning.

Subject

Mx. Johnson claims they have "always loved working with kids" and have always known they wanted to become a teacher since they were young. Because of this choice, Reilly pursued several experiences working with kids in both formal and informal educational settings. Early in their teaching career, Mx. Johnson taught mathematics at a small high school. Reilly transitioned into their current position about two years ago and found the transition to middle school from high school to be "kind of weird." They note that they are "still figuring out what [being a seventh-grade teacher] even means," they "might push [their] kids a little harder because [they] know what they need," and admit they still "have a high school mentality."

At one point, Reilly was going to pursue a career as a music teacher; however, after one year of a music education program, they changed their mind. Instead, Mx. Johnson decided to pursue mathematics teaching "because math is easier" and they missed calculus. However, while they teach mathematics, Reilly values interdisciplinary mathematics as well as discovery-based and project-based learning. As a result, they try to incorporate learning opportunities from other content areas into their mathematics instruction with hands-on activities and projects. Reilly claims that they "don't lecture very often."

Community

As part of their practice, Mx. Johnson interacts with several groups of people both directly and indirectly. Reilly cares for their students as individuals and attempts to create a feeling of individualization when implementing instruction. As part of that instruction, Reilly regularly works with other teachers in the school building to create discovery-based, authentic learning opportunities for students. Indirectly, policies mandated by district and school administration and developed by a team of mathematics leaders as well as parents' views of student performance impact Mx. Johnson's teaching practice.

Students

The course is an accelerated seventh grade course in which Mx. Johnson identifies approximately half of the students as gifted. The students are independent, but "still kids." For example, Reilly claims the students are "extremely energetic and will get very excited when they do well in math." Having experienced standards-based grading for more years than previous students, Mx. Johnson perceives the students as "flexible" and "more okay" with obtaining threes as opposed to fours on a four-point scale, but some students would still say, "I got a two, I failed." Reilly recalls reminding the students that a two means that they are progressing and are on their "way to learning what [they] need to learn."

During classroom instruction, Mx. Johnson expects their students to work together in collaborative teams. This collaboration starts with a discussion of the homework assignment and continues throughout the entire class period. Reilly "rarely has to answer questions" because students are well-normed in using other students and textbook resources to make progress. That is, the students are "self-moderated" and "self-regulated." For example, as students enter the classroom, Reilly expects students to "check their answers, and if they're getting them wrong, they should be figuring out what did [they] do wrong and if [they] can't figure it out, [they'll] ask their team." Only when the team cannot figure out the answer, will Mx. Johnson step in.

To ensure a balanced group, Mx. Johnson uses gradebook data to create ability-based student groups which consist of students with a similar understanding of the current content. By arranging the students in homogenous groups, Reilly believes that they are more able to target instruction to specific students which allows them "to help each [student] more personally, then [they] would as a class."

Depending on student understanding and comfort with the material, Mx. Johnson allows some students to work independently while the whole-class completes an activity. Alternatively, if most of the students seem like they understand the content, Mx. Johnson will pull the few students who are struggling aside to work on remediation while the rest of the class moves onto other tasks and problems. When pulling students aside, Reilly notes that those students are not ready for the content of the lesson and need additional support.

Parents Of Students

Many parents help their child with their schoolwork either personally or by hiring a tutor. Mx. Johnson believes that standards-based grading provides a tool for communication. It allows parents to "know what [their child] did wrong before it gets too far in the year" because parents "really want to know the grades their kids have." However, in Mx. Johnson's view, "parents can't wrap their minds around standards-based grading." Parents "want to equate a four with an A." Reilly notes that "most parents are like 'oh, my kids are not getting fours, they're failing.'" As a result, Mx. Johnson hypothesizes that they spend "a good quarter" of their conferences explaining the differences in grades to parents. Reilly wants parents to have the mindset that their child should be "at a three by the end of the year."

District And School Administration

District and school administrative personnel have an indirect impact on Mx. Johnson's instructional practices and implementation of standards-based grading through, first, the lack of resources and, second, commissioning the development and design of resources. When recalling their introduction to standards-based grading, Reilly notes that there was a week and half of professional development meetings focused on the practice; however, the people leading the professional development expected that Mx. Johnson already knew about the practice. Then, the principal directed them to "have some rubrics ready for the beginning of the year." According to Mx. Johnson, that was the extent of their initial support. Reilly claims they tried to research the idea of standards-based grading, but notes it is "not easy to read up on [and] it's not something that is just innate because [their generation didn't] grow up with it."

Following this first year of struggle, the school district issued a set of rubrics which outlined grade-level specific "competencies" that students must achieve and teachers must assess. A team of "math team leaders" from middle schools within the district created the rubrics. In addition to the rubrics, the school district created and mandated the use of an online gradebook which was divided into weighted categories Reilly called "buckets." Mx. Johnson believes that "the district is the one who determines the buckets" in the gradebook, but that the "buckets" for mathematics come directly from the Colorado Academic Standards for Mathematics [CAS-M] (CDE, 2010). As a result, there are four "buckets" aligned with each of the four CAS-M standards: (a) number sense, (b) algebraic structures, (c) statistics and probability, and (d) geometric relationships.

Other Mathematics Teachers

Mx. Johnson's teaching practice is impacted by two different types of mathematics teachers: (a) a common course teacher, and (b) district-wide mathematics team leaders. Locally, within the middle school, Reilly has the opportunity to co-plan their course with one other seventh grade mathematics teacher. However, in the previous school year, the other teacher also taught sixth and eighth grade in addition to seventh grade. As a result, it is not an established norm for Mx. Johnson and this teacher to co-plan on a regular basis. During data collection, the two teachers teach different types of seventh grade courses with only the one grade-level seventh grade course as overlap. Reilly notes that they and the other teacher "would only be collaborating on that one class." Thus, while Mx. Johnson could co-plan, they choose not to.

At the district-level, a committee of mathematics team leaders impacts Mx. Johnson's teaching practice and implementation of standards-based grading. This committee was responsible for the development of the district-mandated mathematics rubrics that Reilly uses for student assessment. Mx. Johnson believes that this committee consisted of approximately six to eight people who were the "team leads" in every middle school in the district. The other mathematics teachers in the district only met once to discuss the district-mandated rubrics. According to Mx. Johnson, "people are still figuring out" standards-based grading and assessing student competency.

Other Content Area Teachers

As part of their practice, Mx. Johnson engages in a significant amount of collaboration with other content area teachers through interdisciplinary project planning. As a believer in interdisciplinary, cross-curricular learning, Reilly works closely with other teachers; including teachers of English, social studies, and science, to develop learning opportunities and projects for students that cover content in more than just mathematics. As a result, Mx. Johnson's view of standards-based grading is impacted by the views of these other teachers. In particular, Reilly claims that “other teachers in other fields are not as keen to the ideas” of standards-based grading as they are. They hypothesize that this is because standards-based grading is best suited for mathematics; whereas, the other content areas do not have the supports and resources (e.g., well-defined competencies). For example, the English buckets “seem very vague,” and the science and social studies standards are not divided by grade level which made assessing standards difficult.

Instruments

Mx. Johnson feels well-supported by the curriculum and district-mandated rubrics when implementing standards-based grading practices and policies. However, to implement those policies, Reilly needs to supplement and modify those policies to maintain a focus on the students’ current understanding and work habits. As a result, Mx. Johnson creates and utilizes a hierarchy of assessments to support and assess student growth and understanding, and they are strategic with respect to how they enter students’ scores on these assessments into the gradebook. Such assessments include exams, projects, formative assessments, and homework.

Standards-Based Grading Policies

Mx. Johnson “loves” standards-based grading. They believe that the practice helps them better target student strengths and weaknesses with respect to students’ understanding. They claim that, “as a math teacher, [they’re] like, this is the most amazing thing ever because [they] can look at ratios and proportions, expressions, equations, geometry, and be like, this student is really good in all of these, but man, the geometry is killing them.” In contrast, Reilly notes that, with traditional grading system, they would look at student grades and see that the students would have, for example, "an 80% and [they'd be] like, 'I don't know what that [missing] 20% represents.'" Specifically, Reilly believes that standards-based grading allows them to glean more information about students' mathematical understanding from the students' scores and grades.

Mx. Johnson has a general perception of what each score, on a four-point scale, means; however, the requirements for a score depends on the specific competency they are measuring. In Reilly's view, earning a three means that the student is "at grade level" while earning a four means that the student is "exceeding the expectations." More specifically, to earn a four, students must apply, transfer, and/or justify their answers. Mx. Johnson claims that a two means a student is "progressing" and that they are on their "way to learning what [they] need to learn." Teachers could give students "point five grades;" however, Mx. Johnson believes that many other teachers do not do this. Reilly perceives these grades as "in between a one or two, or three or four." By the end of the school year, Mx. Johnson's goal is for each student to be at a three on each competency.

Mx. Johnson replaces students' grades as they retest on competencies. Reilly keeps the previous grade in the gradebook so that they can track the students' progress, but those grades do not impact students' overall grades. When replacing grades, scores on summative assessments replace scores on formative assessments as well as older scores on summative assessments. Mx. Johnson believes that replacing scores is "more fair that way because if [they] really do believe that depth is more important than speed, if [they] really believe that it doesn't matter how long it takes [their students] to get it, then [they] shouldn't be penalizing [the students] for the 15 ones they got, and then the final two fours don't really impact [the students'] scores that much." Reilly hypothesizes that they are the only one in the school that engages in grade replacement. District policies do not officially prohibit or support the practice of grade replacement; however, Mx. Johnson claims that their supervisor knows about their practice and is supportive of the decision.

Mx. Johnson always records the most recent result on a competency even if it is lower than a previous score, and the gradebook only includes the students' most recent grade toward their current grade. As a result, it is possible for a student's grade to decrease, if they do not retain their previous understanding. Mx. Johnson claims that their choice to replace grades with lower grades is "partly because of how the gradebook works…and partly because [they] want to make sure that [students] are retaining the information that they need." Furthermore, by documenting the students' previous and current understanding, Mx. Johnson claims that multiple stakeholders (e.g., Reilly, parents, and administrators) could see where the students started and where they are now with respect to their understanding, and it gives "the students some level of comfort."

District-Issued Resources

While Mx. Johnson claims that they did not receive professional development with respect to standards-based grading, they do utilize several district-mandated resources. These resources include a set of content rubrics, a set of essential student work habits, and an online gradebook.

District-mandated content rubrics. Developed by a committee of mathematics team leaders, the district issued each middle school mathematics teacher a set of content rubrics to use when assessing students' understanding of key mathematics "competencies." According to Mx. Johnson, the committee created the rubrics by "looking at a list of competencies" and then identifying "which standards each competency address[ed] and work[ed] on the phrasing so that it only applied to the grades that [the rubrics were] being applied to." Each competency is a skill related to one of the four standards outlined in the Colorado Academic Standards (CDE, 2010). Reilly views the rubrics as being more conceptual in nature. For example, "to get a four on almost every rubric, it'll say something about an authentic context" and "always analyze or justify." They find the rubrics helpful in supporting instruction and assessment but plan to "student-fy" the rubrics in the future to put them into more student-friendly terms "so that the students can read them without all this weird jargon."

Mx. Johnson claims that they could determine a score on some rubrics with a single question. For example,

> The system of equations question where you have to show [the solution] in multiple ways. Umm, the one is basically, you know, nothing. The two is, you can solve it in one way. The three as you can solve in multiple ways, and the four is that you can solve in multiple ways and justify and analyze something, something.

For this rubric, Mx. Johnson gives students one question and expects them to solve it in two ways. In contrast, Reilly needs multiple questions to assess other competencies. For example, a concept related to probability might require students to create a probability model to earn a two, solve a probability question for a three, and analyze a probability model or use multiply probability models to earn a four. In this case, Mx. Johnson requires the student to earn a two before earning a three or four. That is, if the student answers the level two question incorrectly and answers the level three question correctly, the student is unable to score above a one because that student did not answer the level two question correctly.

Reilly views some of the rubrics as harder than others. For example, to earn a three for “systems of equations,” students have “to solve [a system] in multiple ways, like if you only know how to do substitution, you get a one or a two and that's it. You cannot get a three until you know how to do it two different ways.” Mx. Johnson feels that is “almost a little bit harsh.” In addition, parents have a “hard time” with questions that require a specific method or strategy to earn a three or four because they believe their child should get the credit for getting the correct answer.

District-mandated work habits. The school district created four “work habits” for students to develop during instruction in addition to content knowledge: (a) collaboration, (b) problem solving, (c) communication, and (d) self-agency. Mx. Johnson views the work habits as a guideline for how to engage in the work of learning, and “as a support” for teachers when talking to students and parents. They note that the teachers at their middle school have “tried to go to more of a, like across the school version of grading [the work habits], but it's still in the works, like it's still kind of rusty.” According to Reilly, not all teachers use or assess the work habits because “depending on how [that teacher’s] class runs, it might not always be possible to do all of those things.”

While work habits are not a part of the students’ overall GPA, Mx. Johnson assesses the students’ achievement and performance of those work habits. During the interviews, Reilly does not specify how they assess problem solving or communication, but they do note strategies for assessing collaboration and self-agency. For a collaboration grade, Mx. Johnson occasionally “go[es] around and write[s], like, how well [the student groups are] working as a team.” This allows them to document the students’ “teamwork improving or decreasing.” Mx. Johnson uses homework to measure self-agency. As they complete the homework check, Reilly asks themself, "Did students do it on time? Did [the student] take pride in it? Does it look nice? Did [the student] actually do it?" If the answers to those questions is in the affirmative, then the students earn a four; otherwise, Mx. Johnson gives the students a two, if not completed, and a three, if partially completed.

Online gradebook. The school district issued each teacher an online gradebook to document and organize their content rubrics and work-habit benchmarks. The gradebook are structured into four "buckets" based on the four CAS-M standards: (a) number sense, properties, and operations; (b) patterns, functions, and algebraic structures; (c) data analysis, statistics, and probability; and (d) shape, dimension, and geometric relationships (CDE, 2010). The gradebook averages the scores in each bucket, and then calculates the average of those averaged buckets to determine a final grade. Mx. Johnson is unconcerned about the possibility that some buckets might have more items than others. It is their belief that the gradebook is "an accurate representation of [the students'] grades."

Since the gradebook is online, it allows the teacher, parents, and students to see a student's grade at any time by logging onto the site. As a result, the grades in the gradebook have the potential to help parents identify concepts on which their child needed to work. To support this, Mx. Johnson names the "assignments" in the gradebook as the title/description of the competency it measures. For example, the gradebook included an assignment titled "the students can draw, construct and describe geometrical figures and describe the relationships between them."

Mx. Johnson utilizes a "stamp book system" to supplement the use of the online gradebook as well as to support their policy of grade replacement. For each student, the system consists of a list of the competencies with space for a series of stamps or marks. As a student achieves a score of three or four, Mx. Johnson stamps the corresponding space in the student's stamp book. Reilly only gives stamps for tests and projects, and they are "few and far between." That is, they use a stamp when learning "reach[es] its maximum capacity." When a student achieves three stamps for a competency, Mx. Johnson rewards that student with a "homework pass." A homework pass means that students earn "a free night where [they] don't have to do anything." Reilly claims that the homework passes are "a little treat" for having "mastered" a competency.

Curriculum Support

To support instruction, Mx. Johnson utilizes a reform-based curriculum that is "more of a discovery-based" curriculum. The curriculum is designed in a spiral structure (i.e., previously taught topics appear with new topics) which is a "technique which [Mx. Johnson] has fallen in love with and will definitely carry with [them] if" they ever change textbooks. In addition, the problem-based design allows Reilly to develop a student-driven instructional design in which they feel they "rarely have to lecture." The curriculum also provides Mx. Johnson with access to a test bank of questions that they "pull from, which helps a lot."

When planning for instruction, Mx. Johnson typically goes through the curriculum section by section. However, there are "some certain sections [they] don't like as much." With those lessons, Mx. Johnson creates their own lessons or uses alternative resources. In addition, the curriculum has a lot of hands-on activities.

However, Mx. Johnson notes that those activities are "not always phrased in the best way." They find that "students don't understand the directions and [the tasks] are not broken down enough where [the students] need pictures of what is going on." As a result, Mx. Johnson often "revamps" those activities and turns them into "labs" where students spend a day exploring a mathematical concept in a hands-on, collaborative way. Reilly refers to completing the labs as "in the textbook, but just in disguise."

Summative Assessments

To assess students' current understanding, Mx. Johnson uses two different types of summative assessments: exams and projects. Reilly administers exams regularly which often cover a wide-range of topics; whereas, they assign projects approximately once per quarter that focus on specific mathematical content.

Exams. Over the course of the year, Mx. Johnson anticipates that they will give approximately 11 exams. Reilly separates each exam into multiple sections relating to one of the four CAS-M standards (CDE, 2010), and then each section includes multiple questions that assess multiple competencies. For each competency, a rubric accompanies each question or set of questions so that the students can "see it right then when they're testing if they're curious as to are [they] at the three or not." However, when it comes to writing an exam, Mx. Johnson claims that "you can't write tests until you understand what the test is supposed to look like, like as far as grades go." In that sense, Mx. Johnson admits that their tests do "not reflect the rubrics because [they] are still learning how that is supposed to work."

Reilly allows students to take as much time as they need to complete exams. As a result, exams might take as many as two class periods to complete. Mx. Johnson admits that, earlier in the year, some students bragged that they could cheat by starting the exam one day, noting what was on it, studying that night, and then coming in the next day to finish. As a result, Mx. Johnson modified their expectations for exam completion "to reduce the amount of the grade being impacted by potential cheating." Specifically, they require that students finish a section on the same day. As a result, some students end up with extra time on the first day because there is not enough time to start another section. Reilly does not want the students to "go home and study how to do it, then suddenly, magically know it the next day." If students end up with extra time, Mx. Johnson allows them to work on other work. If students still do not finish in two days, Reilly will "pull [students] in during lunch and pull [students] in during [their elective period]" to finish the exam.

Mx. Johnson builds reassessment into the assessment process by including topics on each exam that they already assessed. However, there are some topics that Reilly views as more important (e.g., linear functions and graphing) and requires students master the concept in a timelier manner. In this case, Mx. Johnson requires students to retest sooner than the next test. As students earn a three or a four on exams, Reilly adds an additional stamp to the student's stamp book. Mx. Johnson views these exams as "more rigid" and, therefore, worthier of a stamp in the stamp book. For example, Reilly restricts students to only a standard calculator for these exams as opposed to their phone calculators.

Projects. In addition to exams, Reilly occasionally assigns a large exploratory, writing project to students. For example, as part of a geometry unit, Mx. Johnson anticipates assigning students an essay focused on similarity and congruence of shapes.

When assigning this essay, Reilly plans to ask students to explore “something like why are all squares [also] rectangles, but [all] rectangles aren’t squares or that sort of thing.” With this essay, Mx. Johnson claims they are introducing the idea of proofs to their students. For past projects, students received both mathematics and English credit for their essays. Due to the size and expectations, Reilly uses projects and essays as another source of summative assessment and, therefore, includes a stamp in a student’s stamp book for achieving a three or four in a competency. However, assessing competencies is more difficult on a project or essay because it is not explicit which components of the project or essay align with a specific competency.

Formative Assessments

Mx. Johnson gives students a formative assessment two or three times per week. They call these formative assessments “standards dips.” Standards dips are structured like a small quiz and consist of one or two questions that students complete individually in 10 or 15 minutes on an index or note card. Reilly only recently started putting the rubrics on standards dips and is considering rewriting the rubrics to describe “smaller skills that then lead up to the full understanding.” This was because “standards dips test skills within a [competency] rather than the whole [competency].” As instruction on a topic or concept progresses, Mx. Johnson increases the difficulty of standards dips. For example, they might make a task more difficult by adding fractions “because [fractions] tend to trip [students] up.”

Mx. Johnson hypothesizes that standards dips measure “probably an equal amount as tests.” However, it is Mx. Johnson's goal “to have exams be the only thing that really count” towards a student's grade. This modification occurs by replacing standards dip grades with exam grades. For example, Mx. Johnson might have four standards dips for adding, subtracting, multiplying, and dividing rational numbers, but after an exam on that topic, Reilly replaces all four grades with the student's exam grade. According to Mx. Johnson, “by the end of the year, the amount of standards dips that actually count towards [a student’s] grade will be less than 20%. [The student’s grade] is mostly tests, projects, and essays.” Due to this eventual replacement, Reilly is less concerned about student cheating on a standards dip because they will assess the student with a later exam.

Reilly claims that standards dips give them “a general idea of the class” and are “really just for [them] to see what [they] still need to teach.” That is, Mx. Johnson notes that the standards dips tell them where they should focus their attention in upcoming lessons. For example, if all their students earn a three on a standards dip, then Mx. Johnson believes that they “don’t really have to hammer [that concept] anymore, [their students] understand that concept.” At that point, that concept becomes independent practice for the students.

While Mx. Johnson perceives significant value in standards dips to their own planning process, they note that standards dips give the potential for students to be “able to learn from their mistakes almost immediately instead of waiting for a test.” It is Reilly’s goal to return standards dips to students within a week so the students could see what they did wrong. Typically, Mx. Johnson includes the student’s score with feedback on graded standards dips. However, sometimes, Reilly will go over how to do a problem right after the students finish it. Other times, Mx. Johnson gives students the standards dips “without a grade on it or even with a grade on it and be like 'okay, we did this a couple of days ago, I want you to fix it now.'" When doing this, Mx. Johnson does not tell students anything other than the score.

Homework

Mx. Johnson assigns daily homework which consists of approximately four to six textbook problems. The purpose of the homework is to allow students to practice new material while also refreshing their understanding of previous material. That is, "the homework is set up so that the first couple of questions are from that day's lesson and then the others are review problems from previous days' [lessons] so they don't forget what they learned at the beginning of the year." The spiraling structure of the curriculum supports the homework format.

According to Mx. Johnson, homework is an opportunity for immediate feedback. The expectation is that, as students check their work against the provided answer key, they should try to figure out what they did wrong by checking in with their group members. Mx. Johnson notes that not all students do that, but that it is the expectation. Once students discuss the homework with their team, Reilly presents any questions students have, after checking their answers, on the board. When it comes to not completing homework or cheating, Mx. Johnson teaches students that there are "natural consequences to [their] actions in that the students are likely fail their assessments."

Reilly records homework as a self-agency work habits grade.

> If they had it neat, complete, and on time, if they had all of those things, they get a four. If it is complete and neat, but it's not on time as they turn it in tomorrow or they turn in later in the class period it will be a three. Same goes for if it's neat and on time, but they didn't show their work. If they don't show their work, they do not get a four.

Since Reilly uses homework to measure a work habit, it does not count towards the students' GPA. Mx. Johnson views the score in the "gradebook as a conversation piece during conference time." Previously, Mx. Johnson "graded every single homework assignment," but their homework grading is "infinitely less" because they hold the students responsible for checking their own work.

Sources Of Feedback

Mx. Johnson designs instruction and classroom norms to provide students with several sources of feedback related to their mathematical understanding. First, Mx. Johnson expects students to engage in conversations with each other to share their understanding as well as critique the reasoning of others. For example, Reilly regularly asks students to share with a partner their understanding of a concept and to ask probing questions of their partners' understanding. Second, when correcting work on either homework or exams, Mx. Johnson has the expectation that students check their work against a provided answer key. If students answer a question incorrectly, the expectation is that they need to seek help from their classmates or Mx. Johnson to improve their understanding. Finally, when returning corrected work, Reilly includes comments and feedback to guide students in improving their understanding.

Sources Of Remediation

Mx. Johnson offers students opportunities for remediation during and outside of regular class time. During class instruction, Mx. Johnson groups students who are scoring a one or two for targeted remediation using small whiteboards to engage students in additional practice. Mx. Johnson believes that this strategy allows them to give students more personal learning opportunities. Mx. Johnson notes that these students would likely not practice the same questions as everyone else in the class. However, the spiraling nature of the curriculum ensures that the missed topics would “keep coming up again.” That is, those students would “get there hopefully.”

Outside of class time, Mx. Johnson hosts a homework club once a week for students to obtain additional support and to finish exams. Reilly considers the time “a free 30 minutes” where Reilly asks students “what do you need from me?” Mx. Johnson is willing to “just sit back and let [students] work” or do “intensive work” with the students. In Reilly's view, many of the students come to homework club because “they just want[ed] to get their homework done.”

Object

A typical day in Mx. Johnson’s class has the same structure. When students enter the classroom, directions are posted on the interactive board for students to get ready for class. These directions direct students to work on a warm-up and check their homework. At the same time, Mx. Johnson moves from group to group checking student homework. After checking homework, Mx. Johnson answers questions at the board or asks students to answer questions at the board. After this discussion, Mx. Johnson transitions into the main lesson for the day. To start the lesson, Reilly asks a student to read the introduction to the section. After addressing student questions, Mx. Johnson turns over the lesson to the group facilitators who lead their group through a series of tasks. At the end of the lesson, Mx. Johnson summarizes the key ideas from the lesson and answers any student questions. A few times a week, Reilly administers a standards dip to assess students’ understanding.

Mx. Johnson is “a fan of discovery-based learning because [they] firmly believe that, if a student discovers it themselves rather [then having the teacher] just standing at a board and telling them why that's true and you better believe me, [then] it's not going to stick." According to Reilly, this instructional design allows students to "take ownership in" their learning. The student-driven, group learning during each lesson provides students with this opportunity.

Summary Of First Observation

At the start of the first observed lesson, students have their homework assignment out on their desks, and Mx. Johnson moves from group to group checking the assignments for completion. As this process unfolds, Reilly reminds the students that they should be discussing the homework with their teammates. If students do not complete their homework, then Reilly reminds them that the expectation is that they work on it during this time. If students use a homework pass as a replacement for completing the assignment, Mx. Johnson expects those students to participate in the discussion of the homework with the group. After the Reilly completes their homework check, Reilly asks a student to present the solution to a problem focused on using two coordinates to determine a linear equation.

Following the discussion of the homework assignment, Mx. Johnson announces that students work on a warm-up task focused on percentages. Mx. Johnson tells the students that the purpose of the task is for Reilly "to determine how to best support [the students] in breaking down problems." Before the students start the problem, Reilly asks a student to summarize the problem. During previous lessons and activities, Mx. Johnson "notice[s] that there [are] a number of times that [students] miss a question because they didn't read it thoroughly enough or they didn't actually answer what the question was asking." In their pre-lesson reflection, Mx. Johnson describes the warm-up problem as a “three-tiered warm-up” which is intended to “determine where that disconnect [is] coming from and [to] come up with a better way of helping them to solve these problems.”

After students completed the warm-up task, Reilly transitions into the main portion of the lesson. The main goal of this lesson is “to really get [the students] to understand the difference between translations, reflections, and rotations” as well as to highlight the ideas of “exact same shape” and congruence. Mx. Johnson starts the discussion by asking a student to describe the previous day’s lesson which included an activity where students moved an image of a key around a computer screen using options called “slide, spin, and flip.” Reilly told the students that day’s lesson, they would learn the vocabulary of “translation, rotation, and reflection.” As review, Reilly asks the students to describe how to move the same key around the screen to a specific location. As the students explains their thinking, Mx. Johnson documents their thinking on the computer. After this quick review, Reilly defines translation, rotation, and reflection on the board.

After assigning new team leaders, Mx. Johnson asks students to work on a task focused on describing how to move the image of a key from one position to another position in a coordinate plane. After students work on the tasks, Reilly asks students to share their solutions. As students share their thinking, Mx. Johnson documents the students’ thinking on the board. While working at the board, Reilly models their expectations for an answer and emphasizes that order does not matter, so student answers might be different.

Following the presentation of this task, Mx. Johnson asks students to move onto the next task which focuses on moving the graph of a triangle around on a coordinate plane. After giving students some time to work, Reilly notes that students appear to be having trouble graphing the triangle. As a result, Mx. Johnson demonstrates how to graph the triangle on the board and then shows students how to complete the problem by translating the triangle to the right four units and up two units, then by reflecting it across the *x*-axis, and finally by rotating it counterclockwise 90 degrees about the point (3, -2). See Figure 14 for a reproduction of this work.

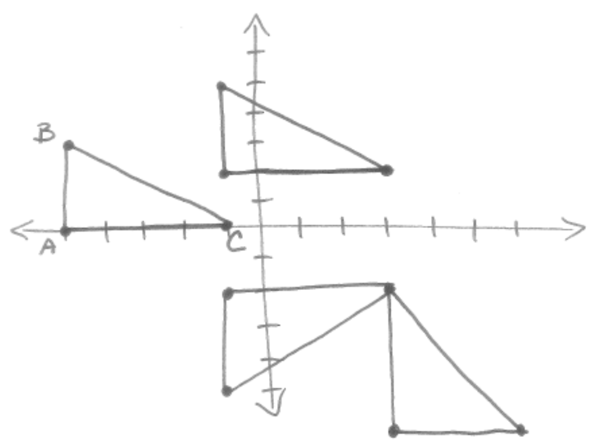

Figure 14. Reproduction of Mx. Johnson's work related to transforming a triangle.

After summarizing the lesson and addressing student questions, Reilly passes out a notecard to each student on which they complete the standards dip. The notecard is special because it has grid lines on it. As Reilly projects the directions for the standards dip at the front of the classroom, students are to (a) clear their desks of everything except for tools, and (b) write their names and the date on the notecards. The standards dip asks students to "find the coordinates of the vertices of each figure after the given transformation. Explain your thinking." See Figure 15 for a reproduction of the problems.

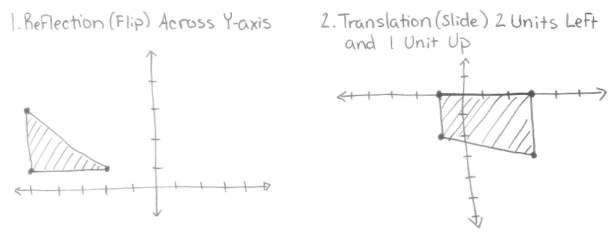

Figure 15. The standards dip problems from the first observation of Mx. Johnson's class.

The standards dip align with competency 5b which states that "the students can draw, construct and describe geometrical figures and describe the relationships between them." According to the rubric projected on the board, to obtain a three, "the student [could] draw, construct and describe geometric figures and cite evidence about the relationships between them." To earn a four, "the student [could] apply and extend their existing knowledge of drawing, constructing, and describing geometric figures to solve and justify familiar and unfamiliar problems in an authentic context." In Mx. Johnson's view, the intent of the standards dip is to determine if students "learned what they were supposed to learn" in the lesson. After they complete the standards dip, Reilly tells students they could leave the classroom for the end of class. Mx. Johnson assigns six textbook problems for homework.

Summary Of Remaining Lessons

The remaining observed lessons focus on transformations of figures in a plane. The second and third lessons continue the first lesson with a focus on rigid transformations by developing vocabulary for describing transformations. The fourth lesson focuses on the implications of multiplication and dilation on figures in the plane. The observed lessons finish with a hands-on lab day that focuses on exploring properties of similar figures.

Second observed lesson. The second observed lesson starts with a homework check and discussion of problems that students request. Several of the questions focus on solving systems of linear equations. For additional practice outside of class, Mx. Johnson tells students to work on "systems of equations" which Reilly assign on Khan Academy as a method for remediation. Following the homework discussion, Reilly announces that, in addition to textbook problems, students would be assigned PARCC practice problems for homework and shows students how to access these problems. Before starting the main lesson, Mx. Johnson returns the previous day's standards dip and notes that they "want [students] to figure out what [they] did wrong." Reilly recommends that, in the future, students should draw the shape and then models the solution on the board.

To start the main portion of the lesson, Reilly asks a student to read the introduction to the new section. Mx. Johnson notes that the lesson is "going to be very self-guided." During small-group time, the student groups work together to complete multiple problems focused on describing changes in coordinate points after performing specific transformations in a plane. Following this work time, Reilly leads the groups in a whole-class discussion of their conjectures and generalization (see Figure 16).

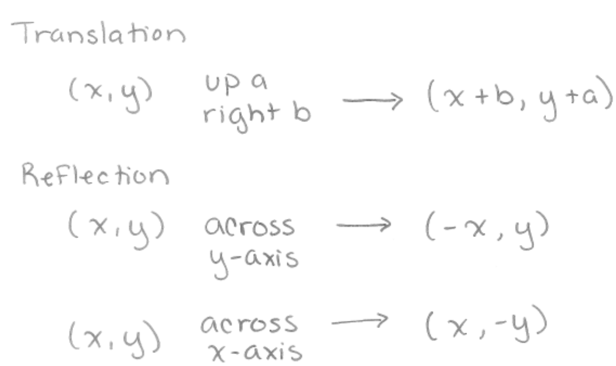

Figure 16. Class generated conjectures about translations and reflections.

At the end of class, Mx. Johnson announces that students are going to complete another standards dip which has the same directions and assesses the same competency as the standards dip from the previous day. The standards dip asks students to "draw each figure and its given transformation. Label the coordinates of the vertices. Explain your thinking." See Figure 17. After the standards dip, Reilly expects the students to complete a reflection answering focused on defining the words translate, rotate, and reflect. The written reflection serves as the students' ticket out the door. For homework, Mx. Johnson assigns three textbook problems and one PARCC practice problem.

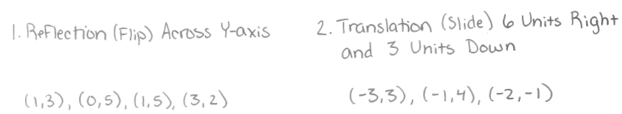

Figure 17. The standards dip problems from the second observation of Mx. Johnson's class.

Third observed lesson. After a check and discussion of homework problems, Mx. Johnson transitions right into the main portion of the lesson. According to Reilly, the "mathematical goal for the day is that [students] should be able to use rigid transformations in context [and] they should be able to understand the difference between reflections, translations, and rotations." Working in their small-groups, students work on an activity which consists of transforming four shapes so that they create the image of a rocket ship. As an extension activity, Reilly ask students to create their own versions of the activity to share with a classmate. Mx. Johnson ends class with a "fun fact Friday" team-building activity. The homework assignment consists of four textbook problems and a PARCC practice problem.

Fourth observed lesson. After a homework check, Mx. Johnson presents a quick summary of the previous lessons that focuses on rigid transformations. As part of this summary, Reilly asks a student to remind the class why rigid transformations are described as rigid. In response to the student's thinking, Mx. Johnson revoices the student's by reminding them that they are rigid because the shapes stay the same shape and size after transforming. As a transition into the main portion of the lesson, Reilly tells the students that the focus of the day's lesson is on "things that do change the shape a little bit." Following this discussion, students work in their small-groups on tasks designed to explore the impacts of multiplying the coordinates of vertices of shapes by various amounts, including values that are negative, greater than one, and less than one, to develop conjectures about the resulting shapes. After students have an opportunity to explore, Mx. Johnson leads the class in a discussion of the students' conjectures and generalizations. The class concludes with an activity during which Reilly asks students to discuss what they learned in class with a partner and to write a reflection focused on how the size of a shape changes when the coordinates of their vertices are multiplied by various values. The reflection serves as the students ticket out of the door. The homework assignment for this class consists of four textbook problems and one PARCC practice problem.

Fifth observed lesson. In the pre-lesson reflection, Mx. Johnson notes this lesson is one of their favorite lessons to teach. The lesson starts "with a homework check as always," but then transitions into a hands-on activity that focuses on similar figures. Reilly claims that they try "to do some hands-on learning at least once or twice a unit." The goal of the activity is for students to investigate "how pantographs change the characteristics of shapes." During the lesson, students use homemade pantographs using rubber bands to explore the "ratios between the side lengths [to determine] how much bigger does a two-band rubber band make a shape and how much bigger does the three-band enlarge a shape." See Figure 18 for a sketch of a student using a rubber band pantograph. Halfway through the lesson, Reilly brings the class together for a discussion of the history of pantographs as well as vocabulary related to similar figures. The lesson concludes with another standards dip assessing competency 5c which states that students "understand congruence and similarity using physical models, transparencies, or geometry software."

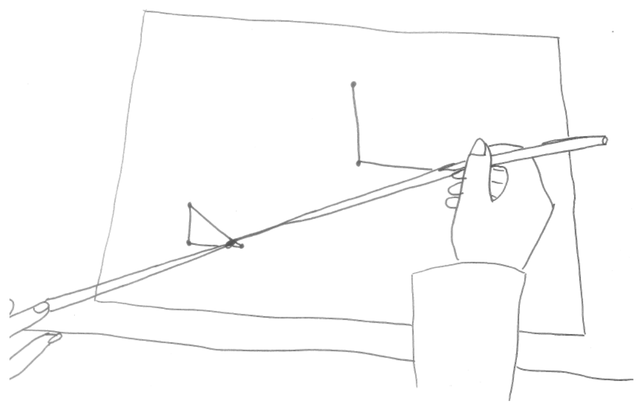

Figure 18. Sketch of a student creating a dilation using a two-band pantograph.

Tensions

While Mx. Johnson claims to "love" standards-based grading, they note several challenges they encounter as part of their experience with implementation. Specifically, Reilly highlights limited professional development and an inconsistent gradebook, as well as a continuing struggle with parent perception of grades as issues impacting their successful implementation of standards-based grading.

Limited Professional Development

After taking their current position, Mx. Johnson claims that they received limited professional development or support on standards-based grading. Reilly recalls feeling as though they were "floundering a lot" their first year. At the beginning of that year, Mx.

Johnson remembers a week and a half of meetings focused on standards-based grading, but the assumption made by the people running the sessions was that teachers were already familiar with the practice. By that point in time, the extent of the information shared during professional development consisted of advice from their principal to "have some rubrics ready for the beginning of the year." Mx. Johnson hypothesizes that it was not until later in their first year that they "finally [understood] what [the session leaders were] talking about and finally was able to explain it to parents." They claim "it took forever" and that was "no excuse for [them to] not know how [their] gradebook works."

During their second year of implementation, Reilly views their current circumstances as improved because they have rough drafts from which to work. They wish that "instead of [coming] up with rubrics [and having] this whole team working on the rubrics for the last two years, [the district] should have waited until those were done or at least had [rough drafts]." To teachers who want to implement standards-based grading, Mx. Johnson recommends starting with rubrics to ensure that there is a common understanding of the possible scores and what they measure. Only then, according to Reilly, can a teacher "worry about how you're going to write the tests." They advise not to "worry about doing it all at once" and to "work through things one at a time."

Changing Gradebook Structure

Mx. Johnson is a supporter of the gradebook structure they are using during data collection. However, over the course of the previous two years, the gradebook had undergone significant changes during the school years. According to Mx. Johnson, the district previously wanted to look at "trending scores," and it was a "total mess." That is, the district wanted "to look at what the students [were] trending at; [were] they increasing? or [were] they decreasing on each standard?" By incorporating trending into the gradebook, the claim was that the students "should go up and down, but eventually they should be going up and down in the threes and fours area." In Reilly's view, "that didn't work."

As a result, in the middle of the previous year the school district changed the gradebook to "an averaging system," which Mx. Johnson views as "not fair" and "totally the opposite of trending." Reilly views trending as promoting a growth mindset in the sense that "trending is hopefully that they're going to get there." In contrast, averaging the students' grades does not make mathematical sense to Mx. Johnson. For example, they note that "if you're averaging those threes and those ones, then you're going to get a two."

When the district made the change, teachers did not lose their grades, but the change in structure modified how grades looked. For example, "kids who were trending high then [they] drop[ped] low, and vice versa." Mx. Johnson recalled that they "had students coming in frantically at the end of the semester" because the students' "grades were all messed up" and the students were "afraid they were not going to pass their classes." Mx. Johnson described this time as "a disaster."

Perceptions Of Grades

When working with parents, Reilly notes two different types of conflicts they encounter. First, many parents have a traditional view of possible grades, and regularly equate a four with an A, a three with a B, and so on. As Mx. Johnson notes, "if you're going to equate it at all, but you shouldn't, a four is supposed to be above and beyond, like above grade level, like really, really exceeding expectations." Instead of talking about student understanding or behavior, Reilly hypothesizes that “a good quarter” of their parent-teacher conferences consist of explaining and justifying the grading scale to parents. In contrast, Mx. Johnson notes that “the kids are fine” and are “more flexible” than the parents.

Second, Mx. Johnson regularly has to explain and justify why a student did not earn a three or four when they produced the correct answer. Reilly recalled a conversation they had with a parent who called and was upset because their child answered a question correctly but did not receive a three. In this case, the student earned a two because the student did not solve the problem using the requested method required by the rubric. While Mx. Johnson views the rubric’s insistence on a method as “a bit harsh,” they followed the rubric and scored the student at the prescribed level two. While the rubrics emphasized the assessment of specific methods and skills, the parent was disgruntled because they felt that obtaining the correct answer should be the goal.

Summary

Mx. Johnson’s case highlights how a teacher incorporated district-developed resources into their implementation of instruction and standards-based grading practices. This case demonstrates how a teacher will interpret and modify district and school policies and practices to support their beliefs about teaching and learning. Reilly’s interpretation of the principles of standards-based grading fit well with their beliefs about discovery- and project-based learning. However, when Mx. Johnson encountered challenges with the school district-issued gradebook, they were able to adapt their practice to fit their practice within the constraints of that gradebook.

CHAPTER VI

MX. ALEX WILLIAMS

When I got rid of unit tests and started assessing on short little standards,
it opened up so much more freedom in my class. It took away the
adversarial relationship between me and my students.
~ Mx. Alex Williams ~

Before students arrive, Mx. Alex Williams passes out notecards, with student names, on desks which are arranged in groups of four around the room. As students enter the classroom, they find their new assigned seat. Music can be heard over the speaker system which creates a welcoming and calming classroom environment. After students are settled with their new group, Mx. Williams projects the plan for the day and transitions the students' attention to the daily lesson starter.

Mx. Williams' case highlights how a teacher took the district-mandated idea of standards-based grading and constructed their own assessible list of concepts as well as strategies for measuring student understanding.

Subject

Mx. Williams describes themself as "very non-traditional" and creative. As a result, their "classroom is not a typical math classroom." For example, they "don't give any homework [and] nothing in class is graded." They were "kind of late to the game" of teaching. Originally, Alex wanted to be a high school physics teacher, but "quickly found out [they] couldn't do any of the math to teach physics." As a result, Mx. Williams "ended up dropping physics and started taking some math classes and kind of fell in love with that and got into math that way." Then, at some point in their college career, Alex took a course in elementary mathematics and was drawn to the challenge of understanding those concepts at a deeper level. Over a decade ago, Mx. Williams completed their degree in elementary education and has taught middle school mathematics ever since.

Mx. Williams started implementing standards-based grading as a personal choice about six years ago. They claim that they were "the only one that" using standards-based grading at that time and they "did it in a traditional grading scale gradebook." When implementing this practice, Alex "had to come up with a way" to make it work in a traditional system. Mx. Williams has taught at their current school for "a couple years" and this school "quote unquote does standards-based grading, so it makes it much easier."

Community

As part of their practice, Mx. Williams perceives themself as working "alone" to develop their practice. Alex feels that it is their responsibility to motive students and communicate student achievement to parents. While Mx. Williams perceives support from school personnel, they claim to experience a lack of support for their practice from other teachers.

Students

The observed class is a "typical" eighth grade mathematics class. Mx. Williams describes the class as their "brightest class" where the students are "all great kids." The students are in "the higher end of [Alex's] spectrum of classes." The expectation is that students "are always working in groups." However, those groups changed every day.

In the past, Alex found that students hated assigned seats which resulted in having to sit by the same person for 10 weeks. As a result, students sit in new seats every day. Mx. Williams assigns the students new seats every day by arranging notecards with their names on them. Alex notes that “most of the time it is random” while other times Mx. Williams is okay with students “working with their friends” or Mx. Williams would create “ability grouping” or “not ability grouping.” Alex “loves that aspect of the cards.” Using the cards also gives Alex a method of quickly taking attendance.

Prior to implementing standards-based grading, Mx. Williams perceived that students “dreaded taking unit tests, and they would do everything in their power to avoid them.” After transitioning to a new system, Alex believes that students no longer perceive them as “giving [the students] a grade.” Instead, students perceive “their grade as tied to their understanding of the standards, and [Alex] is just the teacher that helps them learn these standards.” However, a downside, in Alex's view, is that “too many kids [at their school] are satisfied with twos, like a two is, like, doing pretty good.” As a result, Alex admits that they “started giving a lot more ones, which was hard at first, but it's like [the students are] not, a lot of kids are okay with twos.”

District And School Administration

While Mx. Williams notes that the district has expectations that teachers implement standards-based grading practices, those expectations are not well-defined. For example, Alex notes that “the district has broad categories” for student work habits. Mx. Williams believes that the plan is for teachers to work in the following year to develop “four or five aspects of the work habits that [they're] all going to be on the same page with.” With this development, Alex anticipates that the standards-based grading practices within the district “should be a lot better.” Because of these broad district expectations, Alex feels free to develop their own standards-based grading system and instructional plan within their classroom. With respect to their practice, Alex perceives that their school’s principal “thought it was a great idea.”

Other Teachers

In Alex's view, they are professionally “alone” in their endeavors to implement standards-based grading. While there are “two other eighth grade math teachers in the school, [the group of teachers] don’t really collaborate very well.” Because of Mx. Williams’ decision to not use a textbook, they find it unproductive to work with the other teachers who insist on using a textbook. Alex believes that “a lot of teachers don't want to go away from the book” because “some teachers need [the book] as [their] guidance.” They also believe that those teachers would be resistant to working with Alex because Alex “work[s] too hard.” Mx. Williams recalls conversations with other teachers in which those teachers hypothesized about the amount of time Mx. Williams' strategy would take to implement. In response, Alex notes that they're “like 'dude, you spend way more time grading all those tests.’”

Alex noted that “the hole in [their] system is that like [they're] creating everything.” That is, they were responsible for “what is a level four and what is the level three, [and they're] the one deciding what's the most appropriate skill to assess on.” Alex has “always wanted somebody, at least a partner, that would help develop” their practice. Unfortunately, Alex does not see other teachers adopting their strategies, but thought “it'd be really cool.” They would “love it if [they] could even just get a grade level partner to do it with [them], but [the other teachers] all say, ‘it's too much work.’” In Alex's view, they “just could never see the whole district being able to” develop a common system for implementing standards-based grading.

Parents Of Students

Mx. Williams perceives that, for students, grades were not a motivating factor “unless [the student's] parents are grounding them.” With that in mind, Alex notes that some parents push students to achieve, while other parents do not. Alex recalls that, at parent-teacher conferences, some parents claimed that “[they] didn't graduate high school” and were “fine” in response to Alex noting that their child would need to pass mathematics to graduate. As a result, Mx. Williams views student achievement and motivation as being “mainly up to [themself].”

With respect to their grading system and instructional practices, Mx. Williams claims that “no parent has ever called [them] out on anything and been like, 'well, I don't agree that, that's a level four question versus a level three.'” Likewise, parents have never questioned whether a concept should be assessed. Alex hypothesizes that “being able to retake and everything” appeases parents. That is, Mx. Williams believes that parents view the process as “a very fair system.” However, parents occasionally get upset that Alex did not provide their child with a book.

Instruments

As part of their practice, Mx. Williams developed much of their instruments themself. These self-developed instruments include a list of concepts, a series of assessments, and curriculum resources.

Standards-Based Grading Policies

Alex believes that standards-based grading "changed [their] career; it changed [their] life." Prior to implementing standards-based grading, "kids dreaded taking unit tests and [the students] would do everything in their power to avoid them." Alex recalls that they also "hated" giving and grading unit tests. Alex also disliked the fact that "the unit tests came from the book." By changing their grading and assessment strategies, Alex claims that "it opened up so much more freedom in [their] class; it took away the adversarial relationship between" themself and their students.

Prior to implementing standards-based grading, Alex recalls giving students a "big test" on chapter 4, and a student would come to discuss the C the student received. Mx. Williams admits that the C did not "tell [them] anything because it just meant the [the student] kind of understood chapter 4." To help the student improve or to offer remediation, Alex would "have to go through the unit test and figure out what specific skill or standards [the student] didn't understand." In some cases, the student "could have 90% of all the standards and totally foreign to one." Alex admits that, prior to looking deeper at a student's unit test, they would not have known what the student did not understand. Now, after switching to standards-based grading, they believe that they can look at their gradebook and tell students and parents the concepts on which the student needed to work. For example, Alex might be able to say that the student needs "to learn how to add integers, that is what it holding [the student's] grade back right now." Alex notes that the gradebook serves as "a laser-like remediation tool."

With respect to official, district-mandated policies, Mx. Williams perceives the district rubric as a "loose" description of each level on a four-point scale. Specifically, Alex claims that the district rubric suggests that a "four is exceeding expectations, three is meeting expectations, two is…like getting there, one is not there." However, in Mx. Williams' perception, "there is no expectation" that teachers will use the rubric when determining students' grades. Instead, the rubric is meant to serve as a suggestion for how to interpret student grades. Mx. Williams believes that most teachers are grading by using "a percentage of four…and calling it standards-based grading." For example, a student who scores a 78.5% will receive a 3.14 in the gradebook.

Alex questions what it means to have a grade of four. They note that "some teachers might not say that [a four] is exceeding the standards or maybe it is." In response, Alex asks, "what does it mean to exceed the standard?" Alex notes that many people "equate [a student's grade] with a letter grade" and "even the district makes [the student's grade] a letter grade." Alex describes the grading scale as "an inflated scale." The district gradebook converts the four-point grading scale to a letter grade so that students who average a four will receive an A, a three will equate to a B, and so on. As a result, Mx. Williams notes there are some students who are earning Cs, but in their view those students should be earning Fs. However, as a result, the perception is that a C is "pretty bad." However, parents do not know that Cs are considered bad. In the parent's mind, their student earned a C, so they are "doing pretty good."

As part of their implementation of standards-based grading, Alex notes that, if they are to assess "every skill and every standard, [they] wouldn't be able to do it without just quick[ly] quizzing every single week or even more than that." As a result, Alex does not "assess every single standard because [they] would have too many quizzes." Instead, Alex assesses a self-developed list of "about 30" concepts. When developing their list of concepts, Mx. Williams started by writing down a list of important concepts from the Colorado Academic Standards (CDE, 2010), and then "put them in order of the best flow." While Alex did not use a textbook, the flow of concepts follows a "pretty typical flow from [a] textbook." Then, student grades consist entirely of that student's scores on quizzes covering individual concepts, called concept quizzes. Students' grades are determined by semester. In that sense, Alex has "a semester rolling average" gradebook system. They claim that the "quarter two grade is really a grade for all first semester, and [their] quarter four grade is an average from all second semester."

Alex does not "want the focus of [their] class to be all around grades." They will tell students that "the focus is on the learning and less about the grade." However, by the end of the semester, Alex wants "all [their] kids to have As, and [they] try to make it as easy as possible to get an A." This goal is partially because Alex claims to not care what the student's grade is. What is important to Alex is that their students "enjoy math and become, you know, fall in love with the subject in some ways or appreciate it and love it, and feel successful." Mx. Williams wishes that "there was no grades really, and then [they] could say [to parents], "you could just see all the scores.'"

In addition to grading the students' understanding of concepts, Mx. Williams grades students' work habits. The idea of a work habits grade is a district policy and is new to Alex. Mx. Williams admits that they "don't know how to do that because [Alex wants the work habits grade] to be a holistic grade." In their view, the purpose is "to show parents what kind of students [their children] are." When grading work habits, Alex observes students and considers to what degree are the students modelling good work habits. If the student “generally does them,” they will get a four; where as, if the student “general does not,” then they will get a one.

Concept Quizzes

To assess student achievement of concepts, Mx. Williams administers regular quizzes, called concept quizzes. Each concept quizzes consists of problems assessing three or four concepts. As Alex creates a new concept quiz, they will “add a new [concept] and then take off an old [concept].” As a result, the concepts cycle throughout the year. Concept quizzes occur when Alex feels “like [the students] are ready.” For some concepts, Mx. Williams will give students over a week of practice on a concept before giving an assessment; whereas, with other concepts, Alex assesses after two days of practice. They claim it depends on the concept. On the quiz, each concept is represented by a “level three” question and a “level four” question. See Figure 19 for examples of a level three and a level four “relationships” question.

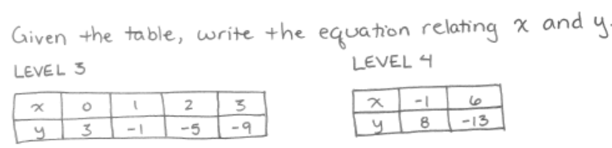

Figure 19. Examples of level three (left) and level four (right) concept quiz questions for the concept “Relationships”.

In the gradebook, a student receives a score for each concept individually, and the gradebook averages each score to get the student’s overall grade in the course. When grading students’ work, Alex claims that they never decrease a student's grade if they do not answer a question correctly. Therefore, it is Alex’s belief that students have “nothing to lose by trying a [level] four” problem after having achieved a score of three. When it comes to reassessment, Alex's belief is that it does not matter “how long it takes [for a student] to learn something as long as [that student] learns it." When a student "learns" the concept by passing a concept quiz, Alex adjusts that student's grade.

When grading a level three question, Alex first looks at whether the student's work is "perfect." If so, then that student will earn a three for that concept in the gradebook. Mx. Williams claims that they can "tell real [sic] fast if [the student] got it right or wrong." If the student's work is "not right," then Alex will "spend [their] time distinguishing between a one and a two." In general, Mx. Williams scores work with "major mistakes" as a one, and work that is "very close" as a two. According to Alex, earning a three on a level three problem means that the student "met grade level expectations."

When it comes to grading a level four task, Mx. Williams' approach differs depending on the students' prior score on that concept. If the student previously earned a three on that concept, then Alex only needs to focus on whether the student answered the questions correctly. If the student answered correctly, then the student receives a four on that concept; otherwise, the student's grade remains a three. If the student had not previously earned a three on that concept, then grading becomes more difficult. If the student answered the questions correctly, then the student receives a four on that concept. If the student answered the questions incorrectly, then the student receives either a two or a one. It is not possible for a student to earn a three based on a response to a level four question. In this case, Alex claims that they would spend their time deciding if the student receives a four, or either a two or one. Earning a three is not an option.

Mx. Williams notes that they will "get kids that want to skip over level three and just do the level four, and then [those students] will get one of [the questions] right, but one of [the questions] wrong." In those situations, Alex struggles to give the student a grade: "did they get a three because they can do one of the level fours? Or do I want to require them to ace level three before they even attempt level four?" Alex decided to put both levels of question on the concept quizzes as the result of frustration by students who would claim that they could do a level three, "so why give [that student] a B when [that student] can get straight to an A."

After assessment, Alex admits that students "definitely lose these skills." However, Mx. Williams' underlying philosophy is "that [the students] are going to forget how to solve a system, but they did it so often and because of the spiraling nature of the quizzes too, that, it won't take long to bring it back." In that sense, the students might lose the skills, but those students should be able to reobtain those skills quicker when they see it again. However, in the future, Alex plans to create specific concept quizzes (e.g., an equations quiz) that they will administer "every quarter because it's just one of those things that [Alex] feels like [the students] need to be able to do all the time."

When developing this method of assessment, Alex claims they did not perceive any "pushback" from parents or students. In fact, Mx. Williams believes that "everybody loved it." They received "some pushback from other math teachers [who] would say it would never work." Alex claims that those other teachers hypothesized that Alex would "have kids come in and take quizzes all the time and do it over and over and over until [those students] got an A without any understanding." The other teachers also thought that Alex's system "was way too much work." Alex admits that "there was a lot of work at first, upfront" developing multiple forms of each concept quiz, but there is little additional work after they developed the assessments because they can reuse those assessments year after year.

Curriculum Resources

Mx. Williams does not teach out of a textbook. They claim that, after five or six years of teaching out of a textbook, the materials were “killing the kids.” Instead they use a book as a resource for themself, but the students never see the book. Because of not using a book, Mx. Williams notes that it caused “a lot of recreating” and “a lot more planning on [their] part.” It is Alex's goal “to be able to give [the students] the resources.” For example, they created a website with instructional videos and practice problems. Unfortunately, Mx. Williams admits that getting students to utilize their website is a challenge. However, Alex recalls some “kids using the website all the time when they feel that they need to get their grade up.” Also, “when [the students'] parents are hounding them, [then Mx. Williams] will see kids watching videos and practicing when they need to come and retake a quiz for the fifth time.” Alex notes that students and parents rarely complain that the students do not have a book, and, when there are complaints, it is “usually when [a student’s] grades are bad.”

One Problems

Over their career, Alex notes that they “usually find the kids that need to be doing the homework are always the kids that don't do the homework [and] those that don't need to do the homework are always the kids that do.” As a result of this perspective, Alex “just quit doing” homework. However, Alex realized that the students were “not studying at all for [the] quizzes, no matter how much [Alex] cheerled [sic] them and provide[d] resources.” In response to this observation, during data collection, Alex started assigning the students “one problem at night” which they refer to as the “one problem.” Mx. Williams assigns the nightly problems “in the hopes that by the end of the week [the students] will do better on the quizzes.” Alex records the students self-recorded completion of the one problem in a record book. Alex anticipates using the students’ self-reported homework checks as part of the students’ work habits grade. Specifically, if the student “did 90% of the one problems they will get a four, if [the student] did 70% to 80%, a three, something like that.”

Feedback

Mx. Williams perceives standards-based grading as more helpful to themself as a feedback and decision-making tool, than it helps students. For example, as a decision-making tool, when Alex looks at the gradebook and sees a student with a really low average, they will then “do some relearning on that.” However, Mx. Williams has “gone back and forth with feedback.” At one point, they would “not even put a grade on [the student's work], like not even a one, two, three, or four.” In these circumstances, Alex would simply highlight students' papers where the students made mistakes. Mx. Williams' perceived that “the kids are so grade driven that [the students] couldn't handle that. It was too stressful.” As a result, Alex started putting grades on students' work. However, in Mx. Williams view, students do not “use the quizzes other than to see, to give them a checkpoint of their understanding on that skill.”

Remediation

Alex claims that their grading strategy allows them “to remediate and differentiate really easily.” For example, “towards the end of the quarter, [they] will bring up grades and will look at all the kids with Ds” and then consider which concept is the “lowest concept and be like, that's where [they're] going to get the most bang for [their] buck." After identifying who those students are, Alex will then, "pull those guys in and tutor them on a certain concept so that they can get their grades up." For example, Mx. Williams might group students so that each group contains students with ones and "a student expert" with a four.

If students need remediation outside of the class period, Alex typically stays after school four days a week. However, Alex notes that none of the students stay after for help. "It's like [the students] don't really care" what their grades are. Instead, Alex approaches students during lunch or in an elective course to encourage them to improve their grades. For example, Alex has "gone up to kids like in band and put little sheets of paper [in front of the students] and be like, 'can you solve that real [sic] quick?.'" According to Alex, these students often do not have a D anymore after having completed a quiz retake. Mx. Williams notes that "not all of [their students] have that sense of ownership in their grade."

According to Mx. Williams, students can also obtain additional instruction on the course website. Alex posted "a bunch of level three problems and a bunch of level four problems so that [students] can click and practice what level they wanted." They also posted "some YouTube video links of people going over the concept." Mx. Williams admits that they try to limit Khan Academy as a resource. Alex notes that, in their mind, the only thing the website was missing is "the written part of books that help tie everything together."

Object

Because the students are eighth graders, Mx. Williams believes "that everything has to be chunked in like 15-minute chunks. If [the lesson] is any longer than that, [Alex] will lose them." As a result, the lesson is segmented into short pieces in which the first half follows essentially the same daily structure. After briefly announcing the outline and plan for the day, students complete one of three "starters": (a) on Mondays and Tuesdays, the starter focuses on a "visual pattern;" (b) on Wednesdays and Thursdays, the starter focuses on estimation; and (c) on Fridays students complete a reflection focused on the learning opportunities from that week. Following a presentation of the solution to the starter, students complete a "quick review" problem focused on a concept they are learning. After giving students time to work, Alex reveals the solution at the board and leads a brief discussion of student errors and confusion. Following the discussion, Mx. Williams shows students a viral video during a portion of class called "this is not math." In Mx. Williams' experience with traditional warm-ups, students will waste "that 5 minutes and not really [do] their warm-up." Instead of wasting time, Alex "started showing a viral video…to bring levity back to the room." During this time of the class period, Mx. Williams returns papers to students. When appropriate, the first half of the class period might also include a discussion of a previous concept quiz or a one problem homework assignment.

While the type of classroom activity changed throughout classroom observations, the final portion of class focuses on presenting and practicing new concepts. During data collection, this portion of the lesson was focused on practicing solving systems using the substitution method, introducing and practicing solving systems using the elimination method, and completing a concept quiz.

Summary Of First Observation

At the start of the lesson, Mx. Williams quickly reviews the plan for the day which includes a starter reflection, quick review, concept quiz review, and a substitution maze (see below). After this quick introduction to the lesson, students spend 5 minutes completing their Friday reflection. The following is the reflection prompt:

> Look back at our estimations and visual patterns for this week. Was there something else you wanted to share on those days but didn't? Whose strategy was new to you that you really liked? In addition to reflecting on the math talks, you can also tell me how math was this week for you. Do you feel you have a good grasp on what we're going? Why or why not?

During the reflection time, Alex reminds students that the expectation is that they spend the entire 5 minutes writing. As students finish the reflection, Mx. Williams collects the students' starters.

After quickly collecting the starters, Mx. Williams transitions the lesson to the daily quick review problem by directing students to get out individual whiteboards in the students' small groups. The quick review problem asks students to solve one of two systems "by substitution." See Figure 20 for a recreation of the two quick review problems. The problem on the left is an example of a level three problem, and the problem on the right is an example of a level four problem. Mx. Williams notes that the difference between the two problems is that the level four problem requires some amount of manipulation before performing a substitution to solve. After students have time to work on a solution to one of the problems, Alex reveals a completed solutions on the board.

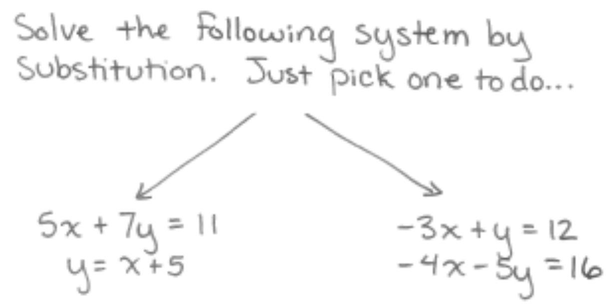

Figure 20. Quick review problem focused on solving systems by substitution from the first observation of Mx. Williams' class.

Following the quick review problem, Mx. Williams moves onto the "this is not math" portion of the lesson. The viral video focuses on a person who trained for an entire year to become a professional table tennis player. The lesson students were to gain from this video is that caring about one's work improves a person's ability to learn as well as learning a little every day is better than trying to learn a lot in a short period of time. While students watch the video, Alex passes back student work which includes a previously completed concept quiz.

Before moving onto the main portion of the lesson, Mx. Williams presents the solutions to the problems from the concept quiz. The concept quiz assesses three different concepts: (a) "slope intercept form;" (b) "relationships;" and (c) "systems by graphing." For each concept, the concept quiz includes several tasks at both level three and level four. See Figure 21 for examples of a level three and level four task, with solutions, from the "slope intercept form" portion of the concept quiz. As Alex presents the solutions at the board, they engage students in a discussion of common errors. For example, on the "slope-intercept form" portion of the concept quiz, some students are confused about the structure of the equation for a line with undefined slope and an *x*-

intercept of 5. Mx. Williams presents a quick example via graphing to justify why the equation for such a line is $x = 5$.

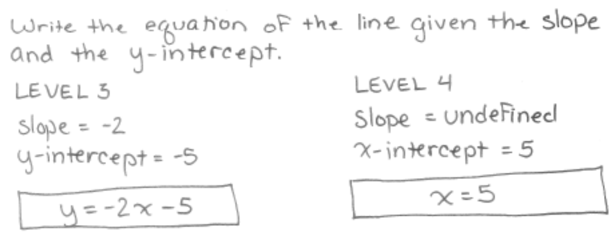

Figure 21. Examples of level three (left) and level four (right) concept quiz questions, with solutions, for the concept "Slope Intercept Form".

During a previous lesson, the class starts working on a "substitution maze" worksheet which asks students to solve systems of linear equations via the substitution method. Based on the answer, the worksheet directs students to move onto another system of linear equations with the goal of reaching the "end" of the maze. See Figure 22 for an excerpt from the substitution maze. To complete the maze, students work on individual whiteboards. While students work, Mx. Williams monitors students and helps as needed. Alex supports students by checking answers, asking guiding questions, or by showing part of the solution process. Students work on the maze until the end of the lesson.

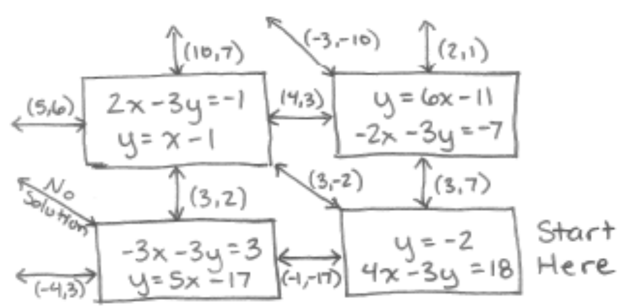

Figure 22. Excerpt from the "Substitution Maze" activity from the first observation of Mx. Williams' class.

Summary Of Remaining Lessons

Due to a school assembly, the researcher was only able to observe Mx. Williams' classroom four total times. During the remaining three lessons, the focus of instruction transitioned towards solving systems of linear equations using the elimination method. The class spent two days developing the need for and practicing this method of solving systems. On the fourth day, the class completed a concept quiz.

Second observed lesson. Since this observation occurs on a Monday, the lesson starts with a "visual pattern" starter. Mx. Williams projects the visual pattern problem on the board and gives students a couple minutes to complete the task. The task asks students to "copy the pattern and draw the next step, create an equation to model the pattern, [and determine] how many objects will be in step 43." Alex emphasizes that they expect students to develop an equation by viewing the pattern visually as opposed to numerically. After students have time to work on the task, Mx. Williams invites a student to share a solution on the board. See Figure 23 for a recreation of the visual pattern and the students' solution. Following the starter, the lesson progresses with a quick review problem focused on practicing solving systems using the substitution method. Mx. Williams announces that the students should expect a concept quiz on the concept by the end of the week.

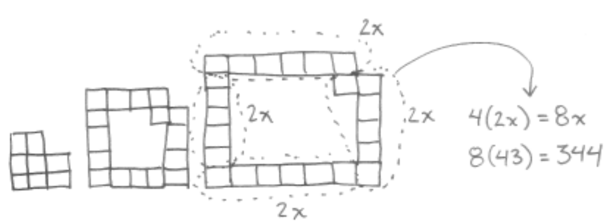

Figure 23. Recreation of the visual pattern starter, with solution, from the second observation of Mx. Williams' class.

Following a quick "this is not math" video, Mx. Williams engages students in an introduction to the elimination method discussion. The discussion starts with an analysis of pros and cons of previous methods including using a table, graphing, and substitution. To motivate the need for the elimination method, Alex asks students to attempt to solve "a trickier" system of linear equations presented in Figure 24. Specifically, Mx. Williams tells students to spend 3 minutes working individually on a solution and then to spend 3 minutes trying "to come to some sort of conclusion with their group." After presenting the solution and new method to the students, Alex asks students to practice the new method by solving two similar systems of linear equations. The lesson concludes with an activity during which Mx. Williams presents systems of linear equations on the board and asks students to identify if they will use substitution or elimination as a solution strategy for the given system. As students prepare to leave the classroom, Alex announces that they are going to start assigning a daily "one problem" and gives each student a copy of the day's one problem which focuses on practicing substitution.

$$\begin{aligned} -7x + 3y &= -12 \\ 7x + 6y &= -24 \end{aligned}$$

Figure 24. Trickier system from the second observation of Mx. Williams' class.

Third observed lesson. To start the lesson, Mx. Williams projects a new visual pattern starter on the board. The directions for the visual pattern starter are the same as the directions for the previous day's visual pattern starter; however, the visual pattern is different. As students work on the starter, Alex asks students to have their completed "one problem" out on their desks so that Alex can record whether the student completed the problem. As announced to students, Mx. Williams plans to "keep track of" the students' completion of the daily one problems so that students will "get one grade in work habits" based on their overall completion over the course of the quarter. Following a quick discussion of the visual pattern, Alex presents the solution to the one problem at the board and transitions to the daily quick review problem. The quick review problem focuses on solving systems by elimination and consists of a level three problem. See Figure 25 for a recreation of the quick review problem. Mx. Williams gives students time to work on the task before revealing the solution on the board.

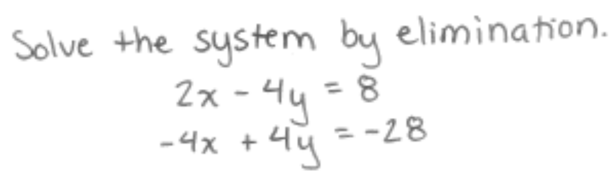

Figure 25. Quick review problem focused on solving systems by elimination from the third observation of Mx. Williams' class.

After a quick "this is not math" video, Mx. Williams reminds students of the challenges of solving systems using tables, graphing and substitution. Alex also notes that there are challenges with solving systems using the elimination method. Specifically, there are systems in which addition does not result in "a variable disappearing" and, therefore, requires a subtraction or multiplication to apply elimination. Following a quick example of such a system, Mx. Williams assigns the students six problems on which to work individually for the remaining portion of class. As students works, Alex monitors their progress and identifies the students' correct answers with a star.

Fourth observed lesson. This lesson starts in the same way as the previous three lessons. The students complete a starter reflection which is followed by a discussion of the "one problem." Following this discussion, the class transitions into completing and discussing a quick review problem which is a level four problem focused on solving systems using the elimination method. Following a quick "this is not math" video, Alex announces that it is time to complete the concept quiz. The concept quiz covers three concepts: (a) "relationships;" (b) "systems by graphing;" and (c) "systems by substitution." Mx. Williams passes out a practice worksheet for students to complete after they complete the concept quiz. The worksheet focuses on practicing level three and level four problems related to solving systems using the elimination method. For the last 15 minutes of class, Alex and their students play a review game which includes problems from a wide range of concepts.

Tensions

When discussing their implementation of standards-based grading, Mx. Williams notes several issues. In particular, they highlight "flaws" they perceive in measuring student understanding, a lack of teacher support in co-planning, and student motivation as challenges to their successful implementation.

Flaws In Measurement

Alex admits that there are "flaws in what [they] do." For example, the questions on the quizzes do not model PARCC exam questions. According to Mx. Williams, the questions on PARCC are "way harder than what [they] quiz on." Since Alex is "quizzing on skills," they note that the argument could be made that they are "not even assessing the standards." In Alex's view, "the standard is to be fluent between different expressions and these deeper questions, [but on their] quizzes [they] don't assess [those] deep concepts." Alex justifies this "flaw" to themself, by noting that, while their quizzes were “not addressing some of the deeper questions, [their] class is structure[d] in a way where [they] are teaching those deeper concepts.”

Alex questions whether students' grades “reflected what [the students] know because they forget things.” For example, Mx. Williams notes that they had a student “accidentally take a level three after [that student had] already gotten a four on a concept and then fail a level three like two weeks later.” Alex admits that other teachers include a retest or retention component to their grading practices, but Alex did not plan to implement such a practice.

Lack Of Co-Developer

Alex believes that getting the rest of the teaching staff on board with a common strategy will “be up for a lot of debate.” However, Mx. Williams admits that “it would be helpful to be able to develop [the system] with somebody.” Alex feels that a co-developer would be beneficial for brainstorming questions and policies. With two other eighth grade mathematics teachers in the school, there is the potential for Alex to have someone with whom to work to develop materials and to brainstorm policies and procedures. However, Alex perceives a lack of support from other teachers when it comes to their instructional practices and standards-based grading practices. In that sense, Alex describes themself as being “alone” with the decision of what concepts to measure and how to best measure those concepts.

Motivation And Perception

According to Mx. Williams, students and their parents exhibit a lack of care and motivation to improve grades. The students are okay with earning a two on a concept and are not motivated to reassess to earn a higher score. Some parents perceive a C as a “fine” score and are not concerned about encouraging their children to improve their scores. Other parents use the argument that they never finished school as an argument for why their child should not have to worry about passing mathematics. Mx. Williams perceives motivating students to achieve and improve as being Alex’s responsibility. However, they note that they might be “enabling” students lack of motivation to improve because they try to never talk about grades in their classroom. This is troubling for Alex because “at the same time, [they] don't want…[the students] to not care about learning the skill either.” Mx. Williams describes this dilemma as “a constant battle.”

Summary

Mx. Williams’ case highlights how, given a district-mandate to standards-based grading, constructs their own assessible list of concepts as well as strategies for measuring student understanding. Alex modified the CAS-M document to create an assessible list of concepts and then created an assessment plan to measure student understanding. Mx. Williams designed and implemented instruction to support their students in successfully completing those assessments.

CHAPTER VII

MX. JAMIE MILLER

The scale that we use is the least important part of standards-based instruction. We became very wrapped up in the trees and lost sight of the forest. In my mind, it's how do we give feedback to students and also their parents about what a student is able to do.
~ Mx. Jamie Miller ~

Prior to the start of class, Mx. Jamie Miller's classroom is alive with the chatter of students who are busying themselves in documenting their homework as well as preparing for the start of the lesson. Jamie can be see preparing for the start of class while also answering student homework questions and engaging students in conversations about the happenings of the day. As students enter the classroom, they take their seats which Jamie has organized into small-groups of four or five students. Jamie has the goal and plan for the day clearly projected on the smartboard at the front of the classroom with the "daily formative" written on the board.

Mx. Miller's case highlights how district-level decisions and values impact a teacher's implementation of standards-based grading. This case demonstrates how the development of resources (e.g., rubrics, online gradebook) as well as the perception of teacher involvement in decision-making can support or hinder the successful implementation of standards-based grading practices.

Subject

Mx. Miller describes themself as a strong mathematics teacher who emphasizes concepts and skills during instruction. Jamie used a lot of mathematics in their former career in finance and finds that mathematical applications align well with their personal interests. Initially, Mx. Miller taught all subject areas as an elementary teacher, but now specializes in mathematics at the middle school level. While they enjoyed the variety that teaching all subject areas provided, they felt that some topics were being "short-changed." Mx. Miller eventually transitioned to teaching middle school with an initial focus on Grade 6 mathematics, but now teaches mathematics at all middle grade levels.

Jamie participated in the school district's initial transition to standards-based grading. While they say "it's not all rosy," Mx. Miller is a strong supporter of many of the practices of standards-based grading, including the collaboration and consistency between and across common course teachers as well as improved communication with students and parents about expectations.

Community

When engaging in the practice of teaching, Mx. Miller interacts with many members of a larger community of people, including students, administrators, a professional learning community (PLC), and other schools within the school district. Each of these groups of people simultaneously support and hinder Mx. Miller's implementation of standards-based grading practices in the classroom by being the focus of instruction (i.e., students) and by supporting and constraining the development of curriculum and policies (e.g., administration).

Students

The students in the observed class are in Grade 6 who are taking an accelerated math course which means that the course included material that is one to two grade levels above their age group. According to Mx. Miller, this class is "a great group of kids" who "are great at working together." The students have "a very high intellectual curiosity, high level of engagement, high level of compliance." Jamie notes that the climate within the class has a "really positive learning focus" where both teacher and students like to laugh and have fun learning mathematics.

District And School Administration

Mx. Miller strongly believes that policies and practices related to standards-based grading need to be teacher-driven to ensure successful implementation. However, according to Jamie, the district-level administration issued many of the current policies using a "top/down approach" which negatively impacted teacher buy-in. For example, the district-level administration issued grading rubrics and a gradebook to all teachers to increase consistency in grades across the district. However, Jamie argues that this goal and motivation has placed too much emphasis on the rubric and scoring while placing little emphasis on the content of learning and assessments within and across teachers.

While Mx. Miller does not approve of the top/down approach to implementation and oversight taken by the district-level administration, they noted that the district has new leadership and they feel as though the district is "on the road to recovery" because the administration seems open to returning some of the decision-making power back to the individual schools. To "fix" the problems with implementation, Jamie notes that "the first thing [the district has] to fix is you gotta [sic] stop making decisions without talking to the teachers first." In addition, they claim that there needs to be enough time to train students, parents, and teachers how to implement and interpret grades before making changes to the system.

In contrast, at their school, Mx. Miller believes they have been well-supported by the principal. During initial implementation, prior to the district's top/down approach to policy and implementation, the principal provided teachers with time to brainstorm, discuss, and develop course materials and policies prior to implementing standards-based grading practices within the school. As part of the developmental period, Mx. Miller was able to work with a PLC to develop rubrics, assessments, and lessons that align with the standards.

Mx. Miller noted that the previous principal was a "very powerful motivator and leader, and he…did a great job of getting teacher buy-in." In particular, this principal allowed the process for teacher meetings to be focused on reaching a consensus about practices and polices related to standards-based teaching and assessment. Jamie viewed the culture as very positive early on where teachers were able to see more of the benefits of standards-based grading. Jamie believes that this has caused teachers to persevere through the current top/down approach from the district. However, Jamie also noted that "you've got to consistently feed the culture if you want to maintain it because it will form on its own and it will be negative…if you're not feeding it good things." With this point, they are noting that teachers who previously saw the benefits can persevere, while other teachers who have less buy-in to the practices might not be able to implement those practices with fidelity. Therefore, Jamie believes it is essential that the administration focus on creating a culture of collaboration and support where teachers feel as though their voices are being heard.

Professional Learning Community

Mx. Miller values a strong PLC community with a focus on student learning as well as action research. Prior to initial implementation, Mx. Miller worked with other mathematics teachers to compare current practices with the goal of developing consistent grading practices from teacher to teacher. Specifically, the teachers met in brainstorming sessions to determine what people were doing with respect to assessment retakes, lesson timing, deadlines, and homework. The group found that teachers were doing very different things even within the same course.

During initial implementation, Mx. Miller was able to work closely with a PLC consisting of common course partners to develop common assessments, rubrics, and curriculum. This collaborative work allowed all teachers of the same course to agree on what students should be able to know at the end of a unit. Jamie believes that providing common course time for teachers to work together is essential for successful implementation of standards-based grading; however, they also claim that you "have to have teachers who are willing to do that." Mx. Miller believes that, absent either one of those, the practice is likely going to fail. Unfortunately, due to multiple course assignments and scheduling, they do not get to work in these teams anymore. As Jamie noted, "if [they] had planning with all of the teachers that [they] teach similar things with, [they] would be planning with other people all the time."

Other Schools

During initial implementation, the school district made the choice "to lead from the middle." That is, they chose to implement standards-based grading at the middle school without implementing it in the elementary and high schools. This mandate remains the current policy within the district. As a result, the school district sponsored professional development, in the form of a motivational presentation by an outside speaker, was only provided to middle school teachers.

Mx. Miller perceives a lot of resistance towards standards-based grading from teachers in other schools within the school district. They believe this resistance comes from the top/down approach that the school district administration and some school administration took towards developing standards-based grading policies and practices.

According to Mx. Miller, there is a problem with communication and buy-in within the district and community because the district does not require teachers at the elementary and high schools to use standards-based grading. Specifically, students enter the middle school without an understanding of the grading practice. By the time students understand the policies and procedures, they are ready to transition to the high school where those teachers are typically grading using a traditional, letter-grade system. Mx. Miller perceives a resistance from high school teachers towards changing their grading practices because high school teachers are focused on calculating GPAs and are less willing to try new things.

Instruments

With respect to their implementation of standards-based grading and instruction, Mx. Miller is supported and constrained by several instruments planning, implementing, and evaluating mathematical instruction. Such instruments include the district-issued policies, rubrics, and gradebook, as well as assessments, online homework, and curriculum resources.

Standards-Based Grading Policies

According to Mx. Miller, the overarching purpose of standards-based grading "is to give more specific feedback to students and parents about specific skills on what, where students are proficient or not." In addition, since standards-based grading is meant to increase consistency in assessment across teachers, it is meant "to get rid of the educational lottery where, you know, depends on what teacher you get assigned to, the quality of your instruction."

Mx. Miller is critical of the school district's implementation of standards-based grading; noting that, at times, it has been "very poor." In particular, Jamie noted that it seems as though the district is "fixing the aircraft as they were flying it." As part of this implementation, Mx. Miller believes there has been too much focus on the grading scale and numbering system. According to Jamie, you can fully implement standards-based practices without changing the scale and numbering system. Specifically, they argued that changing the numbering system caused confusion and masked the goals of the changes. Jamie wishes that people would focus on the big picture benefits of standards-based grading which they believe are more important than any number.

Officially, Mx. Miller follows the Colorado Academic Standards in Mathematics [CAS-M] (CDE, 2010) when it comes to aligning instruction and assessments to the standards. However, on a day-to-day basis, they reference the Common Core State Standards for Mathematics [CCSS-M] (NGA & CCSSO, 2010) because these align with the district-issued gradebook. The decision about the specificity and alignment of the standards was made at the district-level and has changed every year of implementation. According to Jamie, the school district did not consult teachers about these changes.

When assigning standards-based grades, Mx. Miller considers the whole body of evidence related to a student's work focused on a single learning target. With this view, students receive one grade per learning target as opposed to one grade per task or assignment. Since, in their mind, the goal of instruction is for students to learn the mathematics, it is important to give students many opportunities to demonstrate proficiency. However, time is a limiting factor. Philosophically, Jamie does not have a problem allowing students to correct and reassess repeatedly to change their grades; however, there is a practical issue where there is not enough time to be constantly reassessing on concepts. While Jamie does not have a philosophical problem with changing grades if a student shows improvement, they chose to not do so because it is a practical problem. Jamie allows students to retake formative assessments, but not summative assessments.

Mx. Miller believes that standards-based grading improved instruction within their school by providing teachers a tool to use in terms of planning and assessing that forces them to keep track of what students know and are able to do. Jamie believes that teachers are better able to accomplish instruction that was aligned with the standards as well as improve communication of expectations with students. For example, they note that rubrics with a focus on the standards have helped them focus on the important concepts during instruction. These noted benefits align with Mx. Miller's belief that teaching and grading are connected, and "the value of one is diminished if you're not focusing on them being connected."

Standards-based grading has changed the conversations in the classroom away from grades to statements about what students can and cannot do. Mx. Miller notes that they're "able to articulate…what [they]'re striving for a little bit more easily." However, Jamie also notes that it is difficult to determine the overall impact because of all the changes because there has never been enough time to measure whether something is or is not effective.

District-Issued Rubrics

Mx. Miller notes that rubrics are an important tool for helping students know where they are and what they need to improve on as well as an important tool for supporting teachers in reflecting on and implementing instruction. With respect to Mx. Miller's practice, the school-district issued two rubrics as part of the district's mandate that teachers utilize standards-based grading. The first is a "one size fits all" content rubric for assessing mathematical understand, and the second is a "work habits" rubric for assessing student behavior related to the course.

One size fits all content rubric. To measure student understanding of course content, the school district issued a generic 4-point content rubric. Based on this rubric, students can (a) earn a four (Advanced) if they complete each task correctly and can communicate it well, (b) earn a three (Proficient) if they can provide clear evidence that they know it, but might have made a minor error, and (c) earn a two (Partially Proficient) if there is evidence that they kind of know it, but they made multiple errors or have some conceptual misunderstanding. Teachers also award a one or a zero depending on the degree to which students' work suggests that they do not know the material being assessed.

Prior to the district issuing a content rubric, Mx. Miller used a series of more specific rubrics developed with their common course partners. These rubrics directly aligned to the standards at the most detailed, evidence outcome level. A four (Advanced) meant that a student was able to complete tasks beyond grade level, a three (Proficient) meant that a student demonstrated understanding of the grade level skill, and a two meant that a student understood prerequisite material. Mx. Miller notes that the specificity of these rubrics was a benefit when it came to aligning instruction to the assessments and evaluation. However, the rubrics made it difficult to evaluate student understanding because some students would be able to complete tasks at the advanced and proficient levels but would then struggle to complete prerequisite material.

Work habits rubric. In addition to the rubric used to measure content, Mx. Miller evaluates students' work habits using the school district issued work habits rubric. According to the rubric, there are four work habits: (1) I am productive, (2) I am prepared, (3) I participate, and (4) I have positive classroom behavior. According to Mx. Miller, there is not a specific, district mandated method for measuring each of these work habits. Therefore, it is up to Jamie to determine to measure each habit, and they decided to measure each work habit differently. The work habit "I am productive" is measured daily using the students' homework scores. Jamie measures the other three work habits less frequently due to the amount of time it takes to evaluate each student for each habit. Mx. Miller notes that the work habits rubric helps teachers engage students in reflective thinking.

District-Issued Gradebook

To support teachers in reporting student grades, the school district issued and requires the use of an online gradebook. Mx. Miller finds the gradebook to be "glitchy" and difficult to use. In addition, they note that teachers use it in inconsistent ways which causes issues with communication with students and parents. Furthering this communication issue, Jamie believes that parents and students have a hard time understanding the gradebook in general.

Over the past couple of years, there have been several changes to how teachers record the scores for items in the gradebook. Currently, the gradebook allows teachers to record three different types of grades: formative assessments, summative assessments, and work habits scores. Mx. Miller notes that the division of formative versus summative assessments, as well as the separation of work habits from academic performance "is good."

With respect to academic performance, Mx. Miller notes that teachers initially had to attach or link an item to an evidence outcome. However, some felt that this was too specific. As a result, the district changed the alignment requirement to aligning with grade level expectations [GLE]. Currently, Mx. Miller is required to align at the CCSS-M (NGA & CCSSO, 2010) domain level. That is, when entering an item in the gradebook, they need to select one of five domains to align with: (a) ratios and proportional relationships, (b) the number system, (c) expressions and equations, (d) geometry, and (e) statistics and probability. Mx. Miller disagrees with this decision because aligning at the domain level is too generic. They believe that items should be aligned at the GLE level.

After a teacher aligns an item with a domain, the gradebook then averages the domains together. Mx. Miller notes that this creates a problem where a task in one area ends up weighing more in the overall grade than tasks in another area. For example, the tasks associated with a one-month unit on Geometry are weighing the same as the tasks associated with six months of Expressions and Equations. As a result, Mx. Miller admits to misaligning tasks in the wrong standard in an attempt to correct for these weighting issues.

Assessments

Mx. Miller administers two types of assessments: formative and summative. In conversations with students, these assessments are referred to as "formatives" and "summatives," respectively. The purpose of "formatives" is to measure student understanding when they are "forming [their] understanding" and it is "a check in for kids to understand what they need to work on and it's a check in for teacher to understand so it's informing instruction." In contrast, "summatives" provide a "summary at the end of the unit about what did [students] learn."

Formative assessments. Mx. Miller uses several different types of assessments as formative assessments. As noted above, Jamie believes that formative assessments serve two purposes: (1) they are a check-in for students to understand what they need to work on, and (2) they are a check-in for the teacher to understand their students understanding to inform instruction. Such assessments include quizzes and beginning of class warm-ups. In alignment with Jamie's belief that students are still "forming [their] understanding," students can retake formative assessments up until the summative exam focused on the same content. Based on district policy, formative assessments are worth 29% of a student's final grade.

Prior to the start of classroom observations, Mx. Miller had been using a type of formative called a "Weekly Review and Practice" [WRAP] which was a multi-problem assessment that measured several different learning targets and would take most of a class period. Due to the significant amount of time these were taking, Mx. Miller decided to implement a daily warm-up formative assessment format. These warm-ups focus on one learning target per day and allow for the opportunity for individual student work as well as student presentations and discussion.

Summative assessments. Mx. Miller believes that summative assessments provide a summary at the end of a unit about what students learned. As a result, Jamie administers summative assessments at the end of each unit and did not allow students to retake these assessments. Based on district policy, summative assessments are worth 70% of a student's final grade.

Before administering a summative assessment, Mx. Miller provides students with a practice exam that is a modified version of the practice exam provided by the course textbook. Jamie does this as a way to help students prepare for and feel informed about what is going to be on the assessment. Summative assessments focus on and include key learning targets as another strategy to keep students informed about what content is included on the assessment.

When developing assessments, Mx. Miller notes that they would work with other teachers by first breaking down the standard into learning targets and then write the assessments to align. Jamie notes that this process would ideally happen over a long period of time with several rounds of discussion and revision. Since implementing standards-based grading, Mx. Miller believes that their assessments have "changed to be more focused on learning targets and being very specific about testing those learning targets." In that sense, they claim that their "assessments have become tighter." That is, Jamie has become more intentional about what they do and do not assess.

During initial implementation of standards-based grading, Mx. Miller and their common course partners partitioned summative assessments into three sections based on level of content: (1) partially proficient, (2) proficient, and (3) advanced. However, Mx. Miller claims that this structure was difficult to evaluate because students were able to correctly complete the work for proficient and advanced level content but would struggle to complete the content in the partially proficient section of the assessment. Currently, the summative assessments only include questions that focus on specific learning targets.

Homework

Mx. Miller assigns regular, daily homework which is assessed as an "I am productive" work habit. Jamie views homework as a "quick and easy check in" with students to determine their understanding and comfort level with the current material. In Jamie's view, the purpose of homework is to have "a chance to practice" and it should be a place where the teacher can provide feedback. For that reason, Mx. Miller does not believe in grading homework for correctness. However, Mx. Miller does grade each homework assignment based on completeness. Students can earn a four if their homework is on time, a two if its late, and a zero if they do not submit it.

Most homework assignments are completed using an online homework platform, but some are completed on paper. Mx. Miller expects that each homework assignment should take no more than 30 minutes. As a result, they have a 30-minute rule in which students are supposed to stop working on the homework after 30 minutes even if they have not completed it. When students complete their homework online, the system gives students immediate feedback by notifying them if their answer is correct or incorrect. The online homework system not only allows Mx. Miller to see the whole class performance but also the individual issues. Jamie notes that the online system is nice because it provides instantaneous feedback for students, but it can be "a little picky" about the format of answers. If students elect to complete their homework on paper, Mx. Miller gives them the opportunity to check their work at the start of class using a printed answer key.

At the start of class, Mx. Miller expects students to work together in their small groups to figure out questions they have on their homework. Then, Jamie asks students to reflect on their work on the homework. Specifically, Mx. Miller asks students to self-evaluate their understanding using a 4-point scale. If students score themselves as a three or four, then this communicates to Mx. Miller that the students are "in pretty good shape." If students score themselves as a one or two, then this communicates to Mx. Miller that they need to check-in with those students and possibly target additional support for those students. Because the self-evaluations are a way for students to seek additional help and are not recorded, Mx. Miller believes that students are honest when it comes to completing them.

Curriculum Support

Mx. Miller finds their current curriculum to be teacher-friendly with good topics sequencing. However, they also believe that the content presented is very procedurally focused. As a result, they find that they need to supplement with outside resources to find "more meaningful, more meaty" tasks to develop conceptual understanding.

Sources Of Feedback

According Mx. Miller, students receive feedback about their understanding in several ways. First, students receive instantaneous feedback on their homework via the online homework system or they can check their homework against a printed key at the start of class. Following this check of understanding, students can elect to follow-up with Mx. Miller with any questions or concerns about their understanding. Mx. Miller also gives students feedback during whole-class discussions and small-group or individual work. Finally, Mx. Miller comments on assessments with a focus on ways students can improve their work for subsequent assessments.

On exams, Mx. Miller tries to be specific with their comments related to mistakes and misunderstandings. On occasion, usually where there are significant student misunderstandings, students are allowed to correct their exams. Typically, when correcting exams, Miller pairs students so that there is one student who is proficient and one who was not proficient. Mx. Miller does not like to go over problems as a class because they find that some students disengage and tune out the discussion. Jamie believes this occurs because the students do not believe that the discussions apply to them.

Sources Of Differentiation

Mx. Miller notes that the primary source of differentiation is at the course assignment level. That is, at each grade level, there are three different levels of course: remedial/support, on grade level, and accelerated. However, within courses, Mx. Miller tends to differentiate via grouping more advanced or proficient students with struggling and non-proficient students. According to Mx. Miller, this process is mutually beneficial for both students as the struggling student gain additional instruction and the proficient student has the opportunity to share their knowledge which is important for learning.

Previously, Mx. Miller used tiered assignments but found this practice difficult because students would be at different places which made it difficult to check in and have whole-class discussions. With respect to the observed class, Mx. Miller notes that they do not find that they need to differentiate too often because all students are advanced. However, on occasion, Mx. Miller develops activities and tasks which allow students to work at different levels of understanding, for example, the summative, video game review Jamie does on the final day of classroom observations.

Overall, Mx. Miller finds it difficult to differentiate in a mathematics class. They believe it is likely easier to differentiate in a different content area (e.g., history) because you can make the assignment differentiated while instruction stays the same.

Object

The observed lessons focus on a surface area and volume of solids unit. According to Mx. Miller, the current unit is conducive to a lot of variety; however, the curriculum focuses on procedural fluency. That is, “even though it’s formulaic, it’s a good opportunity to focus on concepts… not just formulas and how [we use] the concept to rebuild formulas.” Jamie believes it is important for students to understand how to apply formulas, therefore they plan to use manipulatives during instruction to help students see the mathematics.

A typical lesson in Mx. Miller’s classroom follows the same general outline: (1) a homework check and self-reflection, (2) followed by a warm-up task, and then (3) the lesson. The warm-up and discussion take approximately 50% of the class time and focus on a problem that transitions into the lesson. Then, the lesson includes learning that involves “all of the senses” and starts with a review of the learning target and goal for the day. Jamie scaffolds the notes which follow an "I do, you do" format where they "write on the board and [students write] the same thing in their notebooks."

Summary Of First Observation

As students enter the room at the start of the first observed classes, Mx. Miller has the learning target for the formative assessment projected at the front of the classroom. At the beginning of class, Mx. Miller announces that they are trying a new type of formative assessment called "daily formatives." Instead of taking one day to assess on several learning targets, Jamie plans to give the students a series of tasks focused on one learning target each day. During this conversation, Mx. Miller leads the students in a discussion about the purpose of formative assessments and emphasizes the purpose of providing the teacher and students feedback about where they are in understanding that material. Mx. Miller concludes the discussion addressing the issue of cheating, explaining how that would only give everyone false information. The purpose behind this switch in structure is because the previous formative assessments have been taking a lot of class time to complete and Mx. Miller believes that this new format will make the process more beneficial for everyone.

The formative learning target for the day is "to find the surface area and volume of a cylinder." Prior to this lesson, the learning targets have focused on finding the volume of pyramids and cylinders. The "daily formative" is geared towards assessing the students' understanding of the process of finding the surface area and volume of cylinders. Mx. Miller directs students to get started on the "formative" tasks which are written on the front board. Th two tasks focus on finding the surface area and volume of a cylinder (see Figure 26).

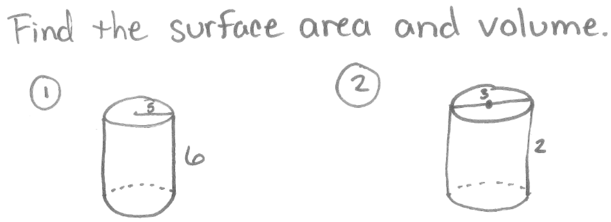

Figure 26. Recreation of the “daily formative” from Mx. Miller’s first observation.

After giving students time to work on these tasks individually, Mx. Miller asks for volunteers to present their solutions on the board. During these presentations, Mx. Miller directs the non-presenting students to reflect on their own work and mark any errors they see with a different color writing utensil. After each student presentation, Mx. Miller asks the non-presenting students to offer two complements for the presenting students. Once students present all four parts of the formative, Mx. Miller directs the students to reflect on and score their own work based on a student-friendly version of the 4-point grading rubric. Specifically, Mx. Miller tells the students:

> So, one through four. You know the meanings of those and you… Think of how I grade, and how I use that number to give you feedback. So, four would be, umm, I could do these in my sleep, I got both of them correct, I have the labels on both of them, I am solid as a rock, give me something harder. A three would be, I’ve got them, maybe I made a calculation mistake, maybe I forgot my labels, but I definitely have the concept and the skill down. Two… And, point fives are okay as well, but I’m not going to describe each of those. Two is that yeah I got some it, like, maybe I got the surface area, but I didn’t get the volume or vice versa, or I made lots of mistakes, or I forgot the formula for the circumference of a circle, something is missing in terms of your overall understanding of it…And then I would like you to give me a reason why you gave yourself that score.

After all students have completed their self-reflection and self-evaluation of their formatives, Mx. Miller shows the students a video clip as a launch into the rest of the lesson. The video shows a person filling a spherical container with water. They then pour that water into a cylindrical container which has the same height and diameter dimensions. The video shows that the water from the sphere only fills the cylinder two-thirds full. Mx. Miller asks the students to reflect on what the two-thirds full result means with respect to finding the volume of a sphere. The purpose of the video is to visually show the students the relationship between the volume of a sphere and the volume of a cylinder. In the past, Jamie found that "students are better able to remember the formula if they understand the concept and to understand the concept, to actually see it is helpful."

After a movement break, Mx. Miller announces that the goal for the lesson is to "find the volume of spheres." The teacher also projects the goal on the board. With the goal of developing the formula for the volume of a sphere, Mx. Miller first asks students to turn to the student sitting next to them and discuss their ideas related to finding the volume of a sphere. After giving them time to discuss, Mx. Miller asks students to share their ideas with the whole class while Mx. Miller documents the students thinking on the board. Following this discussion, Mx. Miller guides students in developing intuition about the equivalence of the height and diameter of a sphere and cylinder by using a tennis ball and plastic cup. Using this intuition, a student shares a word phrase for how to find the volume of a sphere from their knowledge of the volume formula for a cylinder. Mx. Miller then guides students in rewriting the formula for the volume of a sphere.

After the development of the volume formula for a sphere, Mx. Miller asks students to work on a task from a worksheet they had previously started. The problem prompts students to find the volume of scoop of a snow cone in the shape of a hemisphere. Before starting the calculation, Mx. Miller asks the students to think individually and then with a partner on how to modify the formula for the volume of a sphere in order to find the volume of a hemisphere. After documenting the students' thinking on the board, Mx. Miller guides students in finding the answer. While finding the solution, Mx. Miller emphasizes using both the idea as well as the formula when approaching the problem.

Towards the end of class, Mx. Miller asks students to practice finding the volume of a sphere with diameter equal to four. First, students work individually and, then, check with neighboring students to confirm their answers. Finally, Mx. Miller reminds students of the key ideas from the lesson and recommends that they practice finding the volume of a sphere by completing five problems in the online homework system.

Summary Of Remaining Lessons

During the observed lessons, Mx. Miller's focuses instruction on finding the surface area and volume of various solids. Jamie notes that this unit aligns with the geometry standard, but that they have broken the standard into learning targets and each lesson focuses "on a different aspect of that overall standard." According to Mx. Miller, the unit allows for variety in instructional methods. While the curriculum is formulaic, Mx. Miller believes it is important to emphasize how to use geometric concepts to rebuild the formulas as opposed to just memorizing formulas. Mx. Miller plans to use outside resources to supplement the curriculum in order to create more opportunities for conceptual understanding as opposed to procedural fluency. While the first lesson focuses on the development of the formula for the volume of a sphere, the follow-up lessons focus on (a) combining formulas to calculating the volume of composite solids, (b) connecting ideas of similarity, surface area, and volume, and (c) review of unit concepts.

Second observed lesson. The second observed lesson starts with a "daily formative" that assesses the students' understanding of calculating the volume of spheres. Following the "daily formative," Mx. Miller leads students in a discussion of a problem-solving process focused on answering the following questions: (a) What is the problem about? (b) What is each part of the problem asking me to do? (c) What concepts/skills/formulas do I need to know? and (d) How will I approach the problem? During the remaining class period, the students work in small-groups to complete a series of tasks, the "Root Beer Float" activity, related to finding the volume of different shaped glasses (see Figure 27) in order to determine which glass would result in the most root beer.

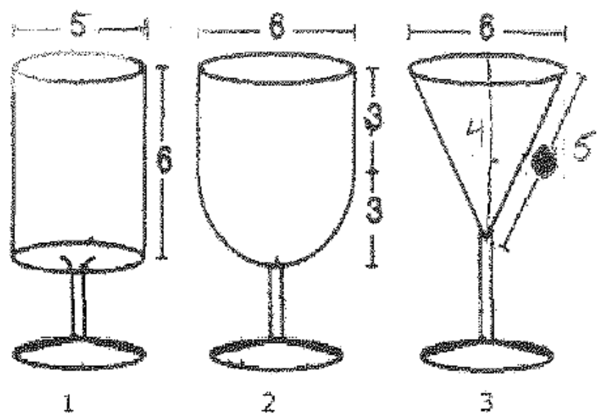

Figure 27. The glasses from the "Root Beer Float" activity.

Third observed lesson. During the third observed lesson, Mx. Miller guides students in an exploration of patterns related to finding the surface area and volume of similar solids. To achieve this goal, the "daily formative" focuses on reviewing key ideas related to proportionality and similarity including calculating a scale factor as well as finding the length of a missing measurement given two similar figures. Following the “daily formative,” Mx. Miller asks students to calculate the surface areas and volumes of similar solids with a focus on identifying possible patterns in their calculations.

According to Mx. Miller, “time was a big factor” in the third lesson. The “daily formative” and homework discussion take longer than expected. As a result, the class is not able to complete the activity related to finding the surface areas and volumes of similar solids. Jamie modifies in the moment the planned homework assignment and asks students to complete the activity as homework.

Fourth observed lesson. The fourth lesson starts with a “daily formative” that assesses students’ understanding of calculating the volume of composite solids. Following the “daily formative,” Mx. Miller leads students in a discussion of the activity from the previous day. Through this discussion, the class develops properties relating the scale factor of similar figures to the scaling of their corresponding surface areas and volumes. At the end of the lesson, Mx. Miller provides the students with a couple practice problems to assess their current understanding of those properties.

Fifth observed lesson. The final observed class focuses on reviewing for the end of unit assessment. The unit assessment assesses the students’ understanding of calculating the surface area and volume of various solids including cylinders and spheres. To assess students’ understanding of the previous day’s lesson, the students complete a “daily formative” that focuses on finding the surface area and volume of similar solids. Following a discussion of the “daily formative,” Mx. Miller asks students to engage in a “video game review” in which students complete tasks at one of 12 “levels” in order to review content from the unit. According to Mx. Miller, this activity “allows students to work at their own pace and reviewing important skills.” The “levels” within the activity start with “the easiest [problems] and then it moves into harder and harder problems.” The activity also offers an opportunity for Mx. Miller to engage with students who are struggling with certain concepts as a form of differentiation.

Tensions

According to Mx. Miller, “there’s huge benefits to standards-based grading.” However, they also note that “it’s not all rosy.” Mx. Miller highlights district implementation, changing the grading scale, and a poorly designed gradebook as obstacles to successful implementation of standards-based grading. In addition, Mx. Miller also identifies difficulties between the philosophical and the practical implementation of the standards-based grading policy of repeated reassessments.

Initial Implementation Resistance

During initial implementation, Mx. Miller notes that many middle schools took a “top/down approach” to implementation in which district and school administration did not include teachers in the discussion of classroom practices and policies. Jamie points to this approach as the reason behind their perception that these “schools are still mostly struggling with the whole idea of standards-based grading” and teachers have limited buy-in to the related practices.

At Mx. Miller’s school, teachers were highly involved with the development of school practices and policies. The work included comparing current practices to develop a consensus and consistency between and across courses with respect to teaching and assessments. Jamie believes that it is this “bottom/up approach” that has result in increased teacher buy-in and perseverance, as well as consistency within and across courses at their school.

Numbering System Changes

As part of initial implementation, Mx. Miller's school did not change the numbering system or scale until district-level administrators mandated the change. Mx. Miller believes that changing the numbering system from a traditional hundred-point scale to a four-point scale "became noise that distracted from the big thing of what [the district was] trying to accomplish." In Jamie's view, the district was initially trying to accomplish "instruction that was more in alignment with the standards and communication that was more clear and detailed." Changing the numbering system created confusion that resulted in resistance from parents and students and lack of buy-in from teachers with respect to standards-based grading. Mx. Miller believes that "you can fully implement standards-based practices without changing the numbering system." If they could go back in time, Jamie would "never" go to a four-point scale; instead, they would go "to a 50 to 100" scale.

Glitchy, Inconsistent Gradebook

Mx. Miller notes that "it's hard to tell overall whether [standards-based grading has] been better for communicating with students because [the school district has] made so many changes in such, like almost every year has been different." One such change has been with respect to the format and structure of the district-issued gradebook. According to Jamie, the "glitchiness and difficulty in use and inconsistency between teachers has made it even harder for parents and students to get on board."

Administrative personal and "tech people" at the district level made the decision about the gradebook design and structure. Over the past several years, the standards alignment within the gradebook has "gone back and forth about the specificity of the standard." Initially, Mx. Miller would align items in the gradebook at the "evidence outcome" level, then the gradebook changed so that tasks were aligned at the "grade level expectation" level. It is important to note that Mx. Miller uses the language of the CAS-M (CDE, 2010); however, the current iteration of the district-issued gradebook uses the language and structure of the CCSS-M (NGA & CCSSO, 2010). The school district personal "made the decision this year to go one level further up." Mx. Miller expressed disapproval of this level of alignment. Specifically, they argued that if the school district is "going to only like to those top-level standards, why bother in the gradebook attaching them to standards because it's too generic."

In addition to the level of alignment required by the gradebook, Mx. Miller expresses frustration with the weighting built into the gradebook. The current iteration of the gradebook creates a weighted score that averages the five overarching standards of the CCSS-M (NGA & CCSSO, 2010). As Jamie points out, this is "very confusing and inaccurate in [their] mind for an overall grade because [the course] may have had one summative on Geometry where we have, we spent six months on Algebra related skills and one month on Geometry, and [the two topics] ended up being weighted the same." Because of this imbalance, Mx. Miller admitted to "putting [tasks] in under the wrong standards because [they] didn't want to create that weighting nightmare. During the time of data collection, Mx. Miller notes that they believe that they gradebook now takes the number of summative assessments into account when determining weighting of the standards.

Managing Reassessments

One of the goals of standards-based grading "is for students to learn," and, as a result, Mx. Miller notes that they "want to give [students] as many opportunities as possible to demonstrate that proficiency." However, "there is a philosophical side of it, and then there's the practical side of it, and those two things don't always go together." Philosophically, Jamie does not have a problem changing students' grades. However, it is practically difficult to constantly change students' grades. That is, it is Jamie's view that it is not manageable for a teacher to allow students to retake every assessment as an open-ended process. As a result, they only allow students to retake formative assessments and not summative assessments. Furthermore, they limit students to retake assessments up to, but not after, a summative assessment on that material.

While students are not allowed to retake summative assessments, if there appears to be significant struggle on a key concept, Mx. Miller will work that concept back into instruction. This reteaching philosophy aligns with Jamie's belief that "sometimes the grading stops, but that doesn't mean the learning has to stop." However, if only one or two students are struggling with that key concept, Mx. Miller will not take time out of whole-class instruction to revisit that concept. Instead, they will "work with [those students] as time allows in the classroom"; however, there "honestly is never enough time to do those one on ones."

Summary

Mx. Miller's case highlights how prior collaborative opportunities help to support and district involvement hinder a teacher's implementation of standards-based grading. As Jamie notes that implementation is "not all rosy," but has benefits when it comes to providing students with improved, purposeful instruction that is more consistent across teachers. However, Mx. Miller argues that the top/down approach to school district policy negatively impacts teacher motivation and willingness to implement those policies with fidelity. As a result, consistent instruction is at risk because teachers lack buy-in to implement the practices.

CHAPTER VIII

CROSS-CASE ANALYSIS

The four teachers who participated in this study were insightful cases of middle school teachers as they worked toward implementing standards-based grading in their practice. Mx. Jamie Miller's case highlighted how district-level decisions and values impacted a teacher's implementation of standards-based grading. Mx. Reilly Johnson's case highlighted how a teacher interpreted and modified district and school policies and practices to support their beliefs about teaching and learning. Mx. Alex Williams' case highlighted how a teacher took the district-mandated idea of standards-based grading and constructed their own assessible list of concepts as well as strategies for measuring student understanding. Finally, Mx. Taylor Brown's case highlighted how an experienced teacher who was new to a school district implemented the district's standards-based grading policies, with little professional development, by accommodating the policies within their personal teaching and assessment strategies, and the constraints of a middle school mathematics classroom. The individual cases and a cross-case analysis provided insights into answers to this study's guiding research questions:

Q1 How do middle school mathematics teachers plan for and structure instruction as part of their implementation of standards-based grading practices?

Q1a What is the nature of the mathematical tasks selected by middle school mathematics teachers for instruction?

Q1b How do middle school mathematics teachers facilitate mathematical instruction?

Q1c What is the nature of classroom discourse as facilitated by middle school mathematics teachers during instruction?

Q1d How do middle school mathematics teachers utilize assessment strategies?

Q2 How do middle school mathematics teachers assign grades as part of their implementation of standards-based grading practices?

Q3 What challenges do middle school mathematics teachers encounter while implementing standards-based grading practices?

Q3a How do middle school mathematics teachers' own teaching philosophies impact their implementation of standards-based grading practices?

Q3b How do other stakeholders (e.g., school district personal, parents, students) impact middle school mathematics teachers' implementation of standards-based grading practices?

Specifically, cross-case analysis of the cases demonstrated similarities and differences between the teachers with respect to their classroom instruction, standards-based grade assignment practices, and challenges faced during implementation of standards-based grading practices.

Instructional Planning And Implementation

The teachers planned for and structured instruction in varied ways as part of their implementation of standards-based grading practices. For each teacher, the researcher analyzed the teachers' first lesson observation to describe the teachers' typical instructional practices. The analysis found similarities and differences related to the teachers' use of mathematical tasks, facilitation of instruction, and classroom discourse during their first lesson observations, as well as the teachers' overall assessment strategies.

Mathematical Tasks

The nature of the mathematical tasks selected by the teachers for use during instruction was one aspect that distinguished the teachers' instructional planning and implementation of standards-based grading. The teachers each used three to seven different mathematical tasks of varying levels of cognitive demand during the first lesson observation (Doyle, 1988; Stein et al., 1996). That is, the teachers engaged students in various tasks whose purpose was to focus the students' attention on mathematical ideas which differed based on the type and depth of thinking required of students to engage with the task. Mx. Brown focused students' attention on varying levels of procedural tasks, Mx. Johnson used a range of tasks including some memorization tasks, Mx. Williams focused on low-cognitive demand procedural tasks, and Mx. Miller used a range of tasks including some doing mathematics tasks.

Mx. Brown's mathematical tasks. During the first observation, Mx. Brown used three sets of mathematical tasks: (a) the four warm-up tasks, (b) the exponents exploration worksheet, and (c) the whiteboard practice activity. As summarized in Table 17 Mx. Brown started with a low-cognitive demand task, followed by a high-cognitive demand task, and then concluded the class period with a low-cognitive demand task.

Table 17

Cognitive demand of tasks used during Mx. Brown's first observation

Task	Cognitive Demand (*Justification*)
Warm-Up Tasks	Procedures Without Connections *Tasks required limited cognitive demand and little ambiguity existed about what needed to be done and how to do it..*
Exponents Exploration Worksheet	Procedures With Connections *Task focused students' attention on the use of a procedure to develop deeper understanding of mathematical concepts and ideas..*
Whiteboard Practice Activity	Procedures Without Connections *Tasks were algorithmic and were focused on producing the correct answers. Explanations focused solely on the procedure that was used.*

Both the warm-up tasks and the whiteboard practice activity were low-cognitive demand, procedures without connections tasks. The warm-up tasks focused students' attention on performing previously learned procedures to analyze and simplify expressions by using properties of exponents and combining like terms. The tasks required limited cognitive demand and little ambiguity exited with respect to what Mx. Brown expected students to do to complete the tasks. The whiteboard practice activity was algorithmic and focused on practicing the previously learned properties of exponents. Taylor expected students to produce the correct answer and student explanations focused on the procedure used to find the answer.

The exponents exploration worksheet was a high-cognitive demand, procedures with connections task. As part of completing this worksheet, Mx. Brown expected students to apply previously learned connections between exponents and repeated multiplication to develop properties of exponents. That is, the goal of the worksheet was to develop deeper understanding of exponents and to develop shortcut rules for exponent expressions.

Over the five observed lessons, Mx. Brown's structure of classroom activity stayed consistent. First, the students worked on warm-up tasks like the one in the first lesson observation. Then the students either worked in small groups or individually developing understanding of the properties of exponents. At the end of class, Taylor led the class in a wrap-up activity which focused on additional practice of simplifying expressions with exponents. Because of this repeated emphasis on the same content over the observed lessons, the cognitive demand of the mathematical tasks reduced to low-cognitive demand as the students increased their understanding of the content. Therefore, while Mx. Brown's first lesson observation included high-cognitive demand tasks, the remaining lessons did not include tasks that achieved high-cognitive demand.

Mx. Johnson's mathematical tasks. During Mx. Johnson's first observation, the students engaged with and completed seven mathematical tasks: (a) homework discussion, (b) three-tiered, percent warm-up problem, (c) key-lock puzzle, (d) definition of key vocabulary, (e) key-lock puzzle with vocabulary, (f) triangle graphical transformation, and (g) standards dip assessment (see Table 18). Reilly's presentation of the definitions of key vocabulary was a low-cognitive demand, memorization task and the standards dip assessment included low-cognitive demand, procedures without connections tasks. The remaining tasks were high-cognitive demand, procedures with connections tasks.

Mx. Johnson's presentation of the definitions of key vocabulary related to transformations of geometric figures focused on committing the definitions of translation, reflection, and rotation to memory without connecting those definitions to the underlying ideas or procedures. Therefore, the task was a low-cognitive demand, memorization task.

The two tasks on the standards dip Mx. Johnson administered at the end of class were low-cognitive demand, procedures without connections tasks. There was little ambiguity about what students needed to do to complete the tasks based on the standards dip's position at the end of a lesson during which the students completed similar tasks. Since the students just started learning the content on the standards dip, the lower level of cognitive demand aligned with Reilly's assessment strategy. Mx. Johnson noted that they tended to ask lower cognitive demand questions on standards dips as students started to learn a concept. Then they increased the cognitive demand of the questions after the students proceeded to develop deeper understanding.

Table 18

Cognitive demand of tasks used during Mx. Johnson's first observation

Task	Cognitive Demand (*Justification*)
Homework Discussion	Procedures With Connections *Task discussed suggested broad, general procedures, and was represented in multiple ways*
Three-Tiered, Percent Warm-Up Problem	Procedures With Connections *Task focused students' attention on the use of procedures for the purpose of developing deeper understanding.*
Key-Lock Puzzle	Procedures With Connections *Task required some cognitive effort. Students needed to engage with the conceptual ideas to complete the task and develop understanding.*
Definition of Key Vocabulary	Memorization *Task involved committing definitions to memory.*
Key-Lock Puzzle with Vocabulary	Procedures With Connections *Task suggested explicit pathways for general procedures that had close connections to underlying conceptual ideas.*
Triangle Graphical Transformation	Procedures With Connections *Task initially required some cognitive effort. Students needed to engage with the conceptual ideas to complete the task and develop understanding. Then Mx. Johnson presented the solution with a focus on using procedures for developing deeper understanding.*
Standards Dip Assessment	Procedures Without Connections *Task required limited cognitive demand and little ambiguity existed about what needed to be done and how to do it.*

The remaining five tasks were high-cognitive demand, procedures with connections tasks categorized into three groups. First, the homework discussion focused on a task that required the interpretation of information using multiple representations to solve the task. The process to solve the task required procedures but merely suggested those procedures using broad, general terms and required students to select an appropriate procedure. Second, the three-tiered, percent warm-up problem, the key-lock puzzle, and the triangle graphical transformation required some cognitive effort and the use of procedures to develop deeper understanding. Third, completing the key-lock puzzle with vocabulary required students to make connections to the previously learned mathematical ideas and vocabulary of translation, reflection, and rotation.

Mx. Williams' mathematical tasks. Mx. Williams used three different mathematical tasks during the first lesson observation: (a) quick review problems, (b) concept quiz discussion, and (c) solving systems with substitution maze. All three tasks utilized by Alex during this lesson observation were low-cognitive demand, procedures without connections (see Table 19).

Table 19

Cognitive demand of tasks used during Mx. Williams' first observation

Task	Cognitive Demand (*Justification*)
Quick Review Problems	Procedures Without Connections *Tasks were algorithmic and did not require explanations beyond the procedure that was used.*
Concept Quiz Discussion	Procedures Without Connections *Tasks were algorithmic and did not require explanations beyond the procedure that was used.*
Solving Systems with Substitution Maze	Procedures Without Connections *Tasks were focused on producing the correct answers and had no connection to the concepts or meaning related to the procedure being used.*

The three tasks taught by Alex during the first lesson were procedures without connection tasks. The quick review problems and concept quiz tasks were algorithmic in nature and did not require explanations beyond the procedure that students used to find the answer. Later, the solving systems with substitution maze focused on producing the correct answer to proceed within the maze. Furthermore, the students' work on the maze did not require any connection to the concepts related to the procedure used to find the answer.

Over the four observed lessons, Mx. Williams' structure of classroom activity stayed consistent. However, during the second lesson observation, Mx. Williams' deviated from this pattern and engaged their students in a high-cognitive demand, doing mathematics task as part of developing the elimination method for solving systems on linear equations. Alex's request for students to develop a method for solving "a trickier" system required students to explore and understand mathematical processes as part of that development. After this exploration, Mx. Williams' presented examples of using the method and provided students with repeated opportunities to practice at a procedures without connections level.

Mx. Miller's mathematical tasks. During their first lesson observation, Mx. Miller asked students to engage with four different sets of mathematical tasks: (a) the two formative tasks, (b) development of the formula for the volume of a sphere, (c) the snow cone hemisphere task, and (d) the calculating the volume of a sphere practice problem. As summarized in Table 20 both the formative tasks and the volume of a sphere practice task were low-cognitive demand, procedures without connections; whereas, the volume of a sphere formula development task and snow cone hemisphere task were both high-cognitive demand tasks, which started as doing mathematics tasks but transitioned to procedures with connections tasks.

The formative tasks and the volume of a sphere practice task were low-cognitive demand, procedures without connections tasks. Both sets of tasks were algorithmic in nature. While the students gave the presentations of the solutions, the formative tasks did not require explanations beyond the procedure the students used to find the solution. The volume of a sphere practice tasks focused on practicing the use of the formula for finding the volume of a sphere to produce the correct answer.

Table 20

Cognitive demand of tasks used during Mx. Miller's first observation

Task	Cognitive Demand (*Justification*)
Formative Tasks	Procedures Without Connections *Tasks were algorithmic and did not require explanations beyond the procedure that was used.*
Volume of a Sphere Formula Development	Doing Mathematics; Procedures With Connections *Task initially required students to explore and understand the nature of mathematical concepts and relationships. Then Mx. Miller focused students' attention on the use of procedures for the purpose of developing deeper understanding.*
Snow Cone Hemisphere Task	Doing Mathematics; Procedures With Connections *Task initially required students to explore and understand the nature of mathematical concepts and relationships. Then Mx. Miller focused students' attention on the use of procedures for the purpose of developing deeper understanding.*
Volume of a Sphere Practice Task	Procedures Without Connections *Task was algorithmic and focused on producing the correct answer instead of developing mathematical understanding.*

Both the volume of a sphere formula development task and the snow cone hemisphere task started as doing mathematics tasks and then, due to implementation, transitioned into procedures with connections tasks. When Mx. Miller asked their students to use their prior knowledge to develop conjectures for formulas for the volume of a sphere and hemisphere, their students were required to explore and understand the nature of mathematical concepts and relationships. Then, when Jamie guided students in refining their formulas, they decreased the cognitive demand by focusing their students' attention on the use of procedures but did so for the purpose of developing deeper understanding. Even considering the reduction in levels, the tasks were high-cognitive demand.

Comparison of mathematical tasks. During the first lesson observation, the mathematical tasks that the four teachers utilized varied in levels of cognitive demand. Only three of the four teachers routinely used high-cognitive demand tasks during mathematical instruction, and the cases demonstrate a spectrum of cognitive demand.

Mxs. Brown, Johnson, and Miller all used high-cognitive demand tasks as part of instruction during their first lesson observation. Mx. Brown used one high-cognitive demand task as the basis for the bulk of the lesson which focused on developing shortcuts for the properties of exponents. In contrast, Mxs. Miller and Johnson used multiple high-cognitive demand tasks. Mx. Miller used several high-cognitive demand tasks to first develop, via doing mathematics (the highest level), and then deepening the students' understanding of the formula for the volume of a sphere by attending to procedures with connections. Mx. Johnson implemented several high-cognitive demand tasks, intermixed with lower cognitive demand tasks, to engage students in developing intuition about transformations of geometric figures. Mx. Johnson's use of low-cognitive demand tasks when presenting vocabulary was consistent with the information the Reilly intended to convey to their students.

If we consider the cognitive demand framework as creating a spectrum of cognitive demand which ranges from memorization to doing mathematics, then the teachers used tasks at varied levels of cognitive demand that demonstrate the spectrum of cognitive demand. See Figure 28 for a representation of the teachers' use of tasks along this spectrum. Mxs. Williams and Brown used tasks which exist in the middle of the cognitive demand spectrum focused on the use of procedures: Mx. Williams utilizing only low-cognitive demand, procedures without connections, and Mx. Brown also using a high-cognitive demand, procedures with connections. In contrast, Mxs. Johnson and Miller utilized tasks at the two extremes of the cognitive demand spectrum: Mx. Johnson focused on the lower extreme, engaging students in memorization tasks as well as focusing on procedures without and with connections, and Mx. Miller focused on the other end of the spectrum, engaging students in doing mathematics tasks as part of their introduction of the volume formulas for a sphere and hemisphere and working on procedures with and without connections.

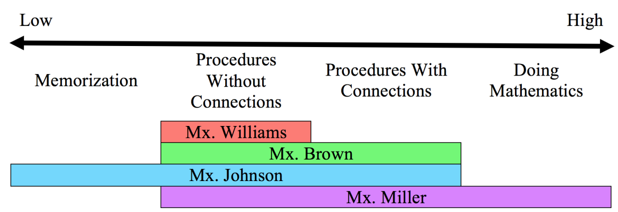

Figure 28. Representation of the spectrum of the levels of cognitive demand of mathematical tasks used during instruction.

Facilitation Of Instruction

The teachers facilitated mathematical instruction, as part of their implementation of standards-based grading practices, using a variety of whole-class, small-group, and individual learning opportunities. Mxs. Brown, Johnson, and Miller used a variety of facilitation techniques while Mx. Williams facilitated instruction primarily through whole-class and individual learning opportunities.

Mx. Brown's facilitation of classroom instruction. As part of Mx. Brown's implementation of the three mathematical tasks during instruction, they utilized all three varieties of facilitation. When completing the warm-up tasks, students worked individually to provide Taylor with a formative assessment of the students understanding of content. Then Mx. Brown led a whole-class discussion focused on the solutions to the warm-up problems. For the exponents exploration worksheet, the students initially worked in small-groups, as pairs, to complete the worksheet. Then Taylor presented the solutions to the worksheet as a whole-class discussion. Finally, the students worked to simplify the whiteboard practice activity tasks individually, and then Mx. Brown presented the solutions as part of whole-class discussions. Overall, Taylor facilitated instruction in such a way that students worked individually or in small-groups to find the solutions to tasks, and then Mx. Brown presented and led a whole-class discussion about those solutions.

Mx. Johnson's facilitation of classroom instruction. While Mx. Johnson utilized all three varieties of facilitation, they primarily utilized small-group and whole-class discussion. Reilly reserved individual work for the percent warm-up problem and standards dip assessment during which they were attempting to assess individual students' understanding of mathematical concepts. During the implementation of three of the seven tasks, Mx. Johnson expected the students to first engage with the tasks in their small-groups and then engage in a whole-class summary discussion of the students' conclusions from the small-group work. Only during the initial key-lock puzzle and the definition of key vocabulary did Reilly engage solely in whole-class discussion. Overall, Mx. Johnson used small-group work to provide opportunities for students to explore and develop intuition about mathematical concepts and then used whole-class discussion to summarize that learning.

Mx. Williams' facilitation of classroom instruction. Mx. Williams utilized both individual and whole-class discussion during classroom instruction. During the quick review problem, the students first completed the problems individually, and then Alex presented the solution on the board in a whole-class presentation structure. After this discussion, Mx. Williams transitioned into a whole-class discussion of the concept quiz solutions. For the remaining portion of class, the students worked individually to complete the solving systems with substitution maze. Overall, Alex allowed their students individual time to practice procedures and then used whole-class discussions to present the solutions to tasks.

Mx. Miller's facilitation of classroom instruction. During classroom instruction, Mx. Miller used all three facilitation techniques. In particular, they used all three varieties of facilitation during instruction for two of the four mathematical tasks used during class. For each task, the students initially worked individually to complete or reflect on the task. Then the students continued to work on the task in some combination of small-group or whole-class discussion; in two tasks, Mx. Miller engaged the students in both. Overall, Mx. Miller used individual learning opportunities so that all students might develop their own ideas before engaging in small-group or whole-class discussions.

For the formative and volume practice tasks, the students initially worked individually to answer the prompts and then worked in either whole-class or small-group formats to discuss their work. Following individual work time on the formative tasks, the teacher facilitated whole-class discussions based on student presentations of the solutions. During the volume of a sphere practice task, the students used their small-group environment to check and discussion the solutions to the tasks.

For both the volume of a sphere and hemisphere formula development tasks, the students worked individually to make sense of and to start the tasks. Following the individual work time, the students used their individual thinking to participate in small-group discussions about their progress which then led to whole-class discussions and development of the formulas and task solutions.

Comparison of instructional facilitation. While Mx. Williams only utilized individual and whole-class instruction, Mxs. Brown, Johnson, and Miller used all three varieties of facilitation: individual, small-group, and whole-class instruction.

The teachers' used individual student work time for two purposes: (a) assessment, and (b) development of understanding. Both Mxs. Brown and Johnson reserved individual student work time to tasks whose purpose was, in part, to assess students' understanding of mathematical concepts. In contrast, Mxs. Williams and Miller used individual work time to develop understanding. In Mx. Williams' case, working alone meant practicing and developing procedural fluency; whereas, Mx. Miller emphasized the development of intuition and conceptual understanding as students worked by themselves.

The three teachers who used small-group instruction did so as a way for students to discuss their conceptual understanding. For Mxs. Brown and Johnson, the small-group learning opportunities accounted for a significant amount of instructional time when students focused on developing conceptual understanding. In Mx. Miller's class, the students briefly worked in small-groups to discuss their ideas as a strategy to formalize their individual thinking before transitioning to a whole-class discussion.

All four teachers used whole-class discussion in their instruction; however, they used them in three different ways: (a) presenting solutions, (b) developing understanding, and (c) summarizing understanding. Both Mxs. Brown and Williams utilized whole-class instruction to present solutions of previously completed tasks and to address student confusion. Mx. Miller utilized whole-class instruction to facilitate the formalization of students' thinking and to develop conceptual understanding. In contrast, Mx. Johnson used whole-class instruction to summarize conjectures and work that students completed during small-group work.

Classroom Discourse

Ellis et al. (2019) argued that classroom discourse had the potential to support and hinder student reasoning and the students' development of mathematical understanding. The four teachers utilized a mixture of discourse moves that had low- and high-levels of potential to support student reasoning in all four areas of the TMSSR framework: (a) eliciting, (b) responding to, (c) facilitating, and (d) extending student reasoning. In addition, during classroom discourse, the teachers offered students a mixture of the four types of verbal feedback: (a) task, (b) process, (c) self-regulation, and (d) self (Hattie & Timperley, 2007).

Mx. Brown's classroom discourse. As part of their first lesson observation, Mx. Brown supported students' reasoning using a variety of discourse moves. During the warm-up task discussion and whiteboard practice activity, Taylor engaged students with primarily low-level eliciting, responding to, and facilitating student reasoning moves. Specifically, the classroom discourse focused on (a) eliciting answers, facts, and procedures, (b) re-voicing and validating student responses, and (c) providing procedural and summary explanations. In contrast, while engaging students with the exponents exploration worksheet, Mx. Brown utilized a mixture of low- and high-level facilitating and high-level extending student reasoning moves. During both small-group and whole-class instruction, the classroom discourse focused on (a) cueing students' attention to reference problems, (b) providing guidance and building on students' prior thinking, (c) providing procedural and summary explanation, and (d) encouraging reasoning and generalization.

During instruction, Mx. Brown offered students feedback focused on the students' current understanding and behavior. With respect to their understanding, when completing the mathematical tasks, Taylor focused their feedback on the students' correct solution to the task and the process the students used to find the solution. In addition, Mx. Brown offered students feedback about behavior. For example, as part of their summary discussions, Mx. Brown offered strategies for taking control of their learning and self-regulation, and, in response to a positive sub-report, Taylor gave their students feedback focused on the students' positive behavior and rewarded the students with no homework.

Mx. Johnson's classroom discourse. Mx. Johnson engaged students with different types of discourse and feedback based on small-group versus whole-class instruction. During whole-class instruction, which was primarily focused on summarizing student learning, Reilly used low- and high-level eliciting, responding to, and facilitating student reasoning moves. In particular, Mx. Johnson engaged their students in discourse by (a) eliciting answers, facts, and procedures, (b) eliciting ideas, (c) re-voicing, (d) re-representing, and (e) providing procedural and summary explanations. In contrast, during small-group instruction, Reilly used mainly high-level facilitating and extending moves which focused on (a) providing guidance and encouraging multiple solution strategies, and (b) encouraging generalization. Furthermore, during both small-group and whole-class discussion, Mx. Johnson uniformly offered students feedback which focused on the students' correct solution to the task as well as the process the students' used when completing the task.

Mx. Williams' classroom discourse. Mx. Williams utilized primarily low-level eliciting, responding to, and facilitating student reasoning moves during the implementation of the three mathematical tasks used in their first observation. In particular, Alex (a) elicited facts and procedures, (b) corrected student errors, and (c) provided procedural explanations and information. As part of working with individual students while the students completed the solving systems by substitution maze, Mx. Williams engaged about a fourth of the students using high-level responding to and facilitating teacher moves by (a) prompting error correction and (b) providing guidance. At the same time, Alex observed other students, but did not directly engage students in discourse. During other lesson observations, Mx. Williams engaged students in similar discourse patterns, but it is unclear from the data collected how Alex chose the students with which they worked nor if they chose to work with different students each day.

Mx. Williams gave students verbal feedback that focused on finding the correct answer and the process used during task completion. As part of the "this is not math" portion of the first lesson observation, Mx. Williams offered students feedback focused on self-regulation and motivation towards learning by suggesting that success is connected to having a positive attitude about learning and developing skill. Specifically, they encouraged students to maintain a positive mindset with respect to learning, and as a result, the students will likely improve their understanding.

Mx. Miller's classroom discourse. Like Mx. Johnson, Mx. Miller utilized different discourse structures depending on the type of task in which they were engaging students. During the discussion of the formative tasks and the volume of a sphere practice task, Jamie used low- and high-level eliciting and high-level extending moves. Initially, Mx. Miller engaged students by (a) eliciting answers, and (b) pressing for explanations. Then they (c) encouraged students to reflect on the process. When completing the development of the formula for the volume of a sphere and the snow-cone hemisphere tasks, Mx. Miller used a variety of discourse strategies including low- and high-level eliciting, responding to, facilitating, and extending student reasoning. In particular, they (a) elicited facts and procedures, (b) elicited ideas, (c) engaged in re-representing and prompting error correction, (d) provided procedural and conceptual explanations, and (e) encouraged evaluation, reasoning, reflection, and generalization. Over the course of the lesson, Jamie provided students with regular feedback related to all four areas of work including task, process, self-regulation, and self. In addition, Mx. Miller asked students to engage in a self-evaluation process (i.e., student-driven feedback) during which students reflected on and wrote their feedback related to the task and process.

Comparison of classroom discourse. All four teachers used a variety of discourse strategies for supporting student reasoning. Mxs. Johnson and Miller mixed low- and high-level discourse moves during every task used during instruction. In contrast, Mx. Brown used low-level moves during all portions of class but also utilized high-level moves when students completed the inquiry-based, exponents exploration worksheet. Mx. Williams, however, used primarily low-level discourse moves as part of the lesson. During individual work time, Alex engaged in high-level discourse with the students with whom they interacted.

With respect to providing students with feedback about their understanding, all four teachers mostly offered feedback to students focused on their correct or incorrect answer to the task and the process they used to find that solution. Except for Mx. Johnson, the other three teachers offered self-regulation feedback once or twice during instruction that focused on strategies for developing effective work habits as well as suggestions for improving understanding. Only Mx. Brown offered students feedback focused on the students' behavior when they praised the students for good behavior for a substitute teacher. Unlike the other teachers, Mx. Miller engaged students in self-evaluations during which the students provided their own feedback related to their understanding of the task and solution process.

Assessment Strategies

During classroom instruction, the teachers utilized various assessment strategies to measure students' understanding of content. The teachers utilized both formative and summative assessments; however, the structure of these assessments differed for each teacher. In addition, the teachers also differed with respect to how they incorporated their assessment strategies as part of classroom instruction. Mxs. Brown and. Miller were both guided by district policy to give both formative and summative assessments, while Mx. Williams implemented only one type of assessment. Mx. Johnson utilized a mix of the other three teachers' strategies.

As per district policy and expectation, both Mxs. Brown and. Miller gave regular formative and summative assessments; however, their actual assessments and uses in grade calculations differed significantly. Both teachers assigned homework: Mx. Brown used online homework as their regular formative assessment, whereas Mx. Miller used homework only as a source for assessing student work habits. Both teachers engaged students in daily warm-ups: Mx. Miller used this formative assessment as part of their grade assignment, whereas Mx. Brown did not it to contribute to students' grades. With respect to summative assessments, Mxs. Brown and Miller gave regular, end-of-unit assessments; however, the structure of the assessments differed between the two teachers. Both Mx. Brown's online and paper-based assessments closely modeled a traditional exam on which students answered questions that measured their procedural fluency. In contrast, on Mx. Miller's assessments, students answered questions which measured varying levels of both procedural fluency and conceptual understanding, and the assessments included the learning target being measured by a specific question. While Mx. Miller only considered end-of-unit assessments as summative assessments, Mx. Brown gave students mid-unit quizzes, end-of-unit exams, and end-of-semester exams which contributed different amounts towards a student's overall grade.

Whereas Mxs. Brown and Miller assessed students in more traditional modes of formative and summative assessments, Mx. Williams administered regular concept quizzes that served as the sole source of graded assessments of student understanding of concepts. Each concept quiz contributed an equal amount to the students' overall grade and measured one of 30 specific concepts identified by Mx. Williams. On the concept quizzes, students chose between completing a "level-three task" or a "level-four task" depending on the score the student wanted to achieve upon completion of the quiz. The tasks used for concept quizzes focused on measuring procedural fluency and mainly focused on the complexity of applying the procedure to distinguish the level.

Mx. Johnson utilized a mixture of all three of the other teachers practices by administering regular small quizzes, called standards dips, as formative assessments and using a variety of larger assessments as summative assessments. Like Mx. Miller, Mx. Johnson used homework completion as a tool for assessing student work habits. Reilly's use of standards dips as a formative assessment was similar to Mx. Williams' concept quizzes in the sense that the assessments primarily focused on students' procedural fluency. In a similar use of Mx. Brown's multiple levels of summative assessments, Mx. Johnson's used end-of-unit tests and projects, but these assessments focused on measuring both procedural fluency and conceptual understanding. For Reilly, the need to assess both types of fluency on their assessments stemmed from the structure and expectation imposed by the district-mandated grading rubrics which emphasized being able to both provide the correct answer as well as explain and justify that answer.

Standards-Based Grade Assignment

After planning for and implementing instruction, the teachers assessed and graded students' understanding of course content, guided by the teacher's implementation of standards-based grading practices. While all four teachers used a four-point grading scale to evaluate and grade students, the teachers differed on how they (a) defined and interpreted that grading scale, (b) supported and allowed students to improve their grade, and (c) communicated and calculated the students' grades. In addition, the teachers expressed varied degrees of frustration with the design and implementation of their district-issued gradebooks.

Interpreting And Using Rubrics

The teachers were given different levels of support by their school districts for evaluating students' understanding using grading scales and rubrics. Mxs. Brown's, Johnson's, and Miller's respective school districts mandated specific grading scales and rubrics to use when measuring students' understanding of course content, while Mx. Williams' school district merely suggested a grading scale and allowed the teachers to decide what to measure with that four-point grading scale.

Mxs. Brown's and Miller's interpretation and use of rubrics. Both Mxs. Brown and Miller utilized a rubric provided by the school district which was a one size fits all, four-point rubric for measuring student understanding. However, Mx. Brown did not use this rubric on a regular basis and, instead, relied on their "mental calibration" of what the levels of the rubric represented in terms of student understanding. With that in mind, Mxs. Brown and Miller had the same general sense of the interpretation of the top two scores on that grading scale. Specifically, a three meant that their students were able to complete grade level content, and a four meant that their students were able to complete content that was beyond grade level. The teachers differed on their interpretations of a score of two. Mx. Brown viewed a two as indicating that a student was confused about the content; whereas, Mx. Miller interpreted a two as indicating that the student could complete prerequisite material.

The students in Mxs. Brown's and Miller's classrooms received grades aligned with differing levels of specificity with respect to the standards. While Mxs. Brown and

Miller were similar in their interpretations of the four-point rubric, the teachers differed on how they used the rubric to determine a student's score. For Mx. Miller, the student's score directly connected to a specific "learning target" which was a specific skill or concept that aligned to a standard. Jamie printed the learning target on the assessment and communicated the learning target to students. As a result, the same assessment might contribute to several grades depending on the number of learning targets being assessed. In contrast, Mx. Brown used the rubric to assess the test as a single score which was aligned at the larger, standard level. As a result, Mx. Miller's students received grades that aligned with specific learning targets, while Mx. Brown's students received grades that aligned with broadly defined standards.

Both Mxs. Brown and Miller utilized an online learning platform as part of regular instruction and assessment. Mx. Miller used the online platform to support online homework assignments which were contributed to the students' work habits grade. Mx. Miller based the students' grade on the level of completeness and timeliness which meant that a student could earn a four, if the assignment was complete and on-time; a two, if the assignment was incomplete or late; and a zero, if the student never completed the assignment. Mx. Brown equated the percentage earned on the online platform to one of the levels of the grading scale. Recall, for example, if students earned a 90% or above, they received a four in the gradebook for that assignment or assessment.

Mx. Johnson's interpretation and use of rubrics. Like Mxs. Brown and Miller, the school district supported Mx. Johnson in their implementation of standards-based grading by providing rubrics and a grading scale. However, unlike the one size fits all rubric used by the other teachers, the district issued specific rubrics for each "competency" which were specific skills and concepts aligned to a standard. As a result, Mx. Johnson was unable to give a general interpretation of what it meant for a student to obtain a score on the four-point grading scale. Instead, Reilly needed to interpret a student's grade using the specific rubric for that competency.

When assessing a student's understanding, Mx. Johnson had to use different assessment strategies based on the specific rubric. In some cases, Reilly was able to determine the student's score on the rubric based on the student's response to a single question. For example, when assessing the student's ability to solve a system of linear equations, Mx. Johnson was able to measure students' understanding using a single question. In other cases, Reilly had to use multiple questions to determine students' scores using the rubric. For example, some of the probability-related rubrics required students to complete similar, but different, tasks to achieve a two, three, or four. As a result, Mx. Johnson had to ask multiple questions to cover the range of tasks required by the rubric. To earn four, in the latter case, the student would need to correctly answer all the questions measured by the rubric.

Mx. Williams' interpretation and use of rubrics. Unlike the other three teachers, Mx. Williams' school district did not issue them a rubric or required list of standards to use when evaluating students' understanding. Instead, Alex created their own list of 30 concepts, based on the CAS-M standards, to use as part of their implementation of standards-based grading. While Mx. Williams noted that the school district provided a general recommendation for the meaning of grades, it was not their impression that the school district expected teachers to use the recommendation as part of their implementation of standards-based grading.

To evaluate their self-created list of concepts, Alex created two levels of tasks aligned with each concept to use when determining students' grades based on the student's ability to correctly answer procedurally focused questions at different levels of difficulty. For each concept quiz, the difference between a score of three and a score of four on Alex's grading scale was the difference between the level of challenge encountered when completing a task. For example, for the concept related to solving "systems by substitution," a level-three task included one equation already solved for a variable; whereas, a level-four task required the students to first solve one equation for a variable before substituting the equivalent expression into the second equation. This grading strategy differed from the other teachers who evaluated based on the students' level of conceptual understanding, such as by explaining their process or thinking.

Reassessment And Remediation

Vatterott (2015) argued that a key component to standards-based grading was the ability for students to obtain remediation, engage in reassessment, and improve or change their grade on a concept or standard. While all four teachers engaged in these practices, they did so in varied ways.

Opportunities for remediation. The four teachers provided their students opportunities for remediation and additional instruction. On occasion, the teachers made those opportunities available outside of regular class and, when classroom constraints allowed, the teachers incorporated remediation as part of regular class time.

Except for Mx. Miller, three teachers offered additional instruction outside of regular class time; however, students did not take advantage of these opportunities in all cases. Both Mxs. Brown and Williams claimed that they made themselves available for students outside of class for additional instruction, and they suggested students referenced outside videos and tutorials for support. While Mx. Brown suggested they were successful in getting students to review online resources, both Mxs. Brown and Williams noted that their students did not appear motivated to obtain remediation prior to reassessment. Mx. Johnson also offered student opportunities outside of regular classroom instruction for remediation by hosting a weekly, lunchtime "homework club" during which many students attended but rarely asked for support beyond checking answers on the current homework assignment.

Mx. Miller was not able to provide direct remediation opportunities outside of regular class time. However, their students were able to obtain additional help from their study hall teacher. Jamie noted that they were often unavailable to their students during the study hall period of the day because they taught a different class at that time. Schedulers placed some students in a mathematics classroom for study hall, and the other students had the ability to approach a different mathematics teacher for assistance during study hall.

All four teachers offered their students a variety of in-class opportunities for remediation on either a regular or occasional basis. Both Mxs. Brown and Johnson incorporated remediation as part of regular instruction by creating opportunities to revisit previously covered concepts as part of instruction. For Mx. Brown, these opportunities included the inclusion of older material on warm-ups and occasional in-class activities; whereas, Mx. Johnson's students regularly worked on previously learned material on their homework assignments. The teachers' stated purposes underlying these opportunities highlight the significant difference between the two practices. Mx. Brown revisited topics to practice and prepare for the PARCC standardized assessment; whereas, Mx. Johnson revisited material to support students' retention of knowledge.

During class, Mxs. Williams and Miller supported students by monitoring their progress as they completed tasks. Mx. Williams regularly engaged students in individual work and helped students as needed. Mx. Miller supported students through homework completion and self-reported reflections. If students suggested that they were struggling to understanding, then Jamie would check in with the student to provide additional instruction. On occasion, both Mxs. Williams and Miller would engage students in activities designed to provided remediation. Mx. Williams structured these activities as whole-class review, Jeopardy-style games; whereas, Mx. Miller created pairs of proficient and non-proficient students for peer mentorship activities. Mx. Miller argued that these activities provided additional instruction for the non-proficient student and provided an opportunity to strengthen the understanding of the proficient student.

Opportunities for reassessment. All four teachers allowed their students to reassess for an improved score; however, the frequency and type of reassessments varied greatly among the teachers. Mx. Williams allowed their students to retake any concept quiz at any time over the course of the entire year for no penalty. In contrast, Mxs. Brown and Miller restricted their students' opportunities to reassess. Mx. Brown allowed students to retake their exam once after taking it the first time. If a student missed the original assessment time, then they would only be able to take the assessment once. Mx. Miller typically only allowed students to retake formative assessments and not summative assessments; furthermore; the students could only reassess up until the summative assessment. Unlike the other teachers, Mx. Johnson required that students reassess on content by regularly including previously assessed material on every exam. As a result, students did not have a choice about when or if they would reassess on a topic but rather were expected to demonstrate their understanding at the time of the exam.

Grade replacement opportunities. The four teachers had different policies regarding opportunities for grade replacement as well as the permanency of grades. Mx. Miller only allowed students to reassess on formative assessments and would give students the highest grade earned on those formative assessments. Since students did not have an opportunity to reassess on the summative assessments, Mx. Miller's students could not improve or change their grade on summative assessments. Mxs. Brown and Williams did allow students to reassess and improve their score. Both teachers took a mastery perspective on student performance and, as a result, only recorded the highest level of mastery and did not attend to retention. That is, if a student struggled to perform on a reassessment at a level previously attained, then the students' grade would stay the same as the previous grade achieved as opposed to decreasing to the new level of performance. In contrast, Mx. Johnson always documented the students most recent score on the reassessment. That is, Reilly's students' scores could decrease if they failed to perform as well or better on a reassessment. Mx. Johnson argued that the process of documenting the students' most recent grade resulted in a grade the reflected the students' current understanding.

Use Of Gradebooks

Each of the teachers utilized a district-issued gradebook to communicate student grades; however, the structure of those gradebooks varied greatly. In addition, except for

Mx. Williams, three teachers had no control over the structure and design of the gradebook.

Mxs. Brown's, Johnson's, and Miller's gradebook design and implementation. In general, the gradebooks of Mxs. Brown, Johnson, and Miller were divided into equally-weighted parts that aligned with the largest category of the corresponding standards document used to support the design of their respective gradebooks. Mxs. Brown and Johnson called the gradebook parts "buckets." Mxs. Brown's and Miller's gradebooks were partitioned into five parts which aligned with the five domains of the CCSS-M standards document. In Mx. Johnson's case, their gradebook was partitioned into four parts which aligned with the four standards of the CAS-M standards document.

After determining in which portion of the gradebook to place a grade, only Mxs. Johnson and Miller entered students' grades in a way that required further refinement of the alignment with the standards documents. Mx. Johnson entered grades using the district-developed competency list as their naming convention. For example, a student might receive a score in the gradebook for an item titled "Competency 5b" which measured the expectation that students be able to "draw, construct and describe geometrical figures and describe the relationships between them." This item would contribute to the students' grade as part of the "Shape, Dimension, and Geometric Relationships" portion of the gradebook. Similarly, Mx. Miller entered grades by naming them based on a previously developed list of learning targets. For example, a student might receive a score in the gradebook for an item labelled as "Students can find the surface area and volume of a cylinder" which would contribute to "Geometry" portion of the gradebook. Jamie developed the list of learning targets with their professional learning community [PLC] prior to the current iteration of school district policies (i.e., the top/down approach to implementation). In this case, Mx. Miller's students might receive a score in the gradebook for an item titled "Students can find the surface area and volume of a cylinder" which would contribute to the students' overall grade as part of the "Geometry" portion of the gradebook. In contrast, Mx. Brown entered students grades by naming them based on the type of task. For example, the students might receive a grade for an item titled "Chapter 16 Test" which focused on assessing the students' understanding of the properties of exponents. As a result, Taylor would align the item with the "Expressions and Equations" portion of the gradebook.

In addition to the naming and standards alignment conventions for items in the gradebook, Mxs. Brown, Johnson, and Miller differentiated between formative, summative, and work habits grades. For Mxs. Brown and Miller, the school district determined and built into the gradebook a distribution of weights related to each of these scores. Specifically, formative assessments accounted for 29%, summative assessments accounted for 70%, and work habits accounted for 1% of a student's overall grade. Mx. Brown chose not to assess work habits due to their perception of the time required and the relatively low impact that the category would have on students' overall grades.

Mx. Johnson differentiated between the types of assessments as part of their personal choices related to their implementation of standards-based grading. Specifically, they chose to count their formative assessments as holding less weight and importance to a student's grade than summative assessments. As a result, they would replace a student's formative assessment grade as soon as that student completed a summative assessment measuring the same competency. Like Mxs. Brown and Miller, the school district required Mx. Johnson to measure work habits; however, the students received a separate work habits grade that did not impact their content grade in the course. By the end of the year, Reilly anticipated that their students' overall content-related grades would consist almost entirely of summative assessment scores.

Mx. Williams' gradebook design and implementation. Mx. Williams' school district issued teachers an online gradebook; however, unlike the other three teachers, Mx. Williams' had some control over the structure of the gradebook. To calculate overall grades, the gradebook calculated averages of the items in the gradebook and converted the scores into letter grades. Since Mx. Williams had control over the structure they used entering items in the gradebook, they did not partition the gradebook into parts but, instead, entered 30 equally-weighted items that directly connected to the list of self-developed concepts they assessed using the concept quizzes. For example, a student might receive a grade for an item titled "systems by substitution." The students overall grade consisted of an average of the grades for each concept quiz. The school district suggested that teachers assess students' work habits; however, Alex was unsure how to do that effectively. As a result, Mx. Williams did not assess students work habits as part of a formal grade. They were under the impression that the school was going to work on developing a common list of work habits and a rubric they might use during the following school year.

Implementation Challenges

As part of their implementation of standards-based grading practices, the teachers encountered a range of challenges related to their personal teaching philosophies and the community in which they worked. On one hand, the teachers had to accommodate the new practices with their own experiences and personal teaching philosophies. On the other hand, the teachers had to navigate and incorporate the influences of stakeholders into their practices. In addition to the anticipated challenges with respect to philosophy and community, the teachers also encountered challenges with issued gradebooks. This latter challenge, although common among the four teachers, was unanticipated as a research question.

Teaching Philosophies

The teachers' beliefs about teaching and learning influenced their interpretations and implementation of standards-based grading practices as part of their overall teaching practices. Specifically, their beliefs influenced how and what they assessed as well as how they used their resources to support their instruction. Mxs. Johnson, Miller, and Williams were supporters of standards-based grading practices, while Mx. Brown did not have strong feelings for or against the practices.

Both Mxs. Johnson and Miller valued both procedural fluency and conceptual understanding; therefore, they taught and assessed student understanding of both. During instruction, both teachers used high-cognitive demand tasks to develop intuition and conceptual understanding to motivate the development of a procedure. Then, during assessment, the teachers included tasks that measured both procedural and conceptual fluency. The school district supported Mx. Johnson in this type of implementation by providing content specific rubrics, while Mx. Miller used previously developed learning targets and curriculum alignments from their early implementation PLC work.

While Mx. Williams' claimed to value deep, conceptual learning, their assessments focused on procedural fluency with varied degrees of complexity. The self-developed list of concepts reflected the content from the standards that they viewed as essential skills for students to understand and be able to do upon completion of the course. Even though Mx. Williams identified a need for conceptual understanding, the concepts in the list were rephrased versions of the standards and focused on procedural fluency. By extension, their instruction mirrored that focus on procedural fluency with the purpose of preparing students for success on the concept quizzes which also emphasized procedural fluency.

Mx. Brown held a mixture of reform and traditional beliefs about teaching and learning mathematics. During instruction, they utilized inquiry-based activities to support the students' development of understanding related to the properties of exponents. However, Taylor assessed and emphasized procedural fluency and structured their assessment practices in traditional quiz and exam formats. Mx. Brown used the textbook to determine and support their instruction instead of using a standards document as their guide.

Community

As part of their implementation of standards-based grading practices, the teachers in this study encountered several groups of people who had an interest and concern about the success and fidelity of that implementation. As stakeholders, the values and beliefs of the teachers' respective school districts, coworkers (i.e., other teachers), students, and the students' parents impacted the decisions the teachers were able to make.

The school district. While each teacher's school district mandated them to implement standards-based grading, only Mx. Williams was able to implement the practices in a way in which they were not constrained by district-developed policies and resources. This meant that Alex had to spend a significant amount of time developing their own materials and assessment practices. Specifically, they created their own list of assessible concepts and the corresponding assessments to measure the students' understanding of those concepts.

In contrast to Mx. Williams, the remaining three teachers were all given rubrics and structured gradebooks which determined what and how the teachers planned for instruction and assessed student understanding. Mx. Brown was able to accommodate the rubrics and gradebook into their already developed views of teaching and learning. Even though Mxs. Johnson and Miller felt that the rubrics supported their development of instruction and assessments, they still felt constrained at times. Mx. Johnson felt that the specific rubrics were too restrictive with respect to the mathematics the rubrics expected students to demonstrate; whereas, Mx. Miller felt that the district's expectation that they align at the overarching standard level was too generic. As a result, Mx. Johnson advocated for more broadly defined rubrics, while Mx. Miller advocated for more specificity. Taken together, the teachers recommended a rubric design that strikes a balance between their respective designs.

Other teachers. As part of their implementations, the teachers had differing experiences when it came to support from other teachers and the degree to which they were able to collaborate with their peers. Mx. Miller valued a PLC culture and believed that their work with other teachers during the initial implementation greatly impacted and supported their grading practices. Mx. Miller also had the opportunity to co-plan with other teachers, but felt restricted in that ability because of the different courses they taught and the need to co-plan with too many groups of teachers. Finally, Mx. Williams felt isolated in their practice because other teachers felt that Alex worked too hard. As a result, Mx. Williams felt "alone" during their development and implementation of standards-based grading. Both Mxs. Brown and Johnson had the opportunity to co-plan with other teachers but chose not to due to perceived differences in student populations as well as time constraints.

Students and parents. For the most part, the teachers found their students to be adaptable and in favor of the implementation of standards-based grading practices; however, in some instances, students and their parents were frustrated and confused by the new practice. This confusion caused teachers to spend additional time communicating information about the grading practices. For example, Mx. Johnson hypothesized that they spent a quarter of their parent-teacher conferences explaining to parents how to interpret grades. Overall, confusion and beliefs about the meaning of grades impacted interpretations of grades and student motivation to improve their grade.

All four teachers found that some students and their parents struggled to interpret their grades correctly. Specifically, some students and parents thought of the rubric scores directly converting to letter grades so that a four was an A, a three was a B, and so on. In contrast, the teachers viewed rubric scores as descriptions of student understanding such that a four meant advanced understanding, a three meant proficient understanding, and so on. On one hand, some of students and parents viewed a rubric score of two as a "good" score because it meant they earned a C and were passing, while the teachers' interpretation was that such a student was failing. On the other hand, some students and parents were unhappy with a score of three because this meant they had earned a B, but they expected to earn all As, while the teachers' interpretation was that the students were proficient and working at grade level.

In addition to issues with interpretation, Mxs. Brown and Williams believed that grades were not a motivating factor for their students. By extension, because the students did not view grades as motivating, they were not motivated to improve their grades through remediation and reassessment. Mx. Brown claimed that they had a very low number of their students who reassessed to improve their grades, and Mx. Williams believed that motivating their students was "mainly up to [them]." Both teachers acknowledged that student motivation was one of the arguments for the using of standards-based grading (Deddeh et al., 2010); however, neither teacher experienced this benefit as part of their implementation. This lack of student motivation to improve their understanding and grade was not discouraging for Mx. Williams and their implementation of standards-based grading; however, it did negatively impact Mx. Brown's perceived value in the practice.

Gradebook

An unanticipated challenge to implementing standards-based grading were the challenges the teachers expressed with respect to their use of the district-issued gradebook. The teachers expressed varied amounts of frustration and concern related to the design and usefulness of their district-issued gradebooks. The teachers expressed issues relating to the gradebook's predetermined weighted averaging and student perceptions of grades. To work around these issues, some teachers developed adaptations to their practice or the gradebook to make the gradebook more usable.

Gradebook weighting. When it came to calculating the students' overall course grades, the teachers questioned if the resulting grade was representative of the students' understanding of course content. Mxs. Brown, Johnson, and Miller noted that the students' overall grade was impacted by disproportionately weighted partitions of the gradebook based on the number of items contributing to each partition. For example, Mx. Miller noted that their students, due to the algebraic focus of the course, might only have a small number of items in the "Geometry" portion of the gradebook, while the students might have a large number of items in the "Expressions and Equations" portion of the gradebook. If a student performed poorly on the geometry-related tasks and performed well on the other tasks, then the student's grade would be lower than what Mx. Miller perceived the student's grade should be. Both Mxs. Brown and Johnson acknowledged the issue but felt that the students grades still worked out okay in the end.

Perceptions of grade conversions. Mx. Williams did not support the school district's conversion of the student's grades into letter grades. They questioned if the students' grades based on the four-point grading scale meant the same as the socially understood interpretations of the corresponding letter grades. For example, a student who averaged a two received a converted letter grade of C as their overall grade in the gradebook. Alex interpreted a student who was earning a C in their course as failing the course; whereas, Alex perceived students and parents as interpreting a C as being a passing grade. As a result, Mx. Williams perceived the letter grade conversion as creating confusion about the meaning of students' grades.

Gradebook adaptations. Both Mxs. Brown and Johnson admitted to using and developing strategies to make the gradebook reflect their personal standards-based grading practices. For example, Mx. Brown recalled instances in which they would enter the same item in the gradebook more than once to artificially increase that items overall impact and weight in the gradebook. Other teachers and administration recommended this process as a strategy to allow Taylor to account for the different levels of assessment (e.g., mid-unit quiz versus end-of-chapter test) used in their classroom. Mx. Johnson questioned the school district's focus on documenting a student's trending scores, and instead decided to only include the student's most recent scores in the overall grade calculation. To accomplish this modification, Mx. Johnson used a paper-based recording system (i.e., a stamp book) and the online gradebook's "do not include in grade calculation" function to document and modify students' overall grades.

Summary

The previous sections highlighted the unstandardized nature of the teachers' implementation of standards-based grading. The teachers used varied levels of cognitive demand mathematical tasks during instruction supported by varied types of instruction and moves used to engage students in supporting student reasoning while completing those tasks. When assessing student understanding, the teachers all administered assessments, but the type (i.e., summative versus formative) and mathematical focus (i.e., procedural versus conceptual) of those assessment varied with respect to frequency and overall impact on student grades. After assessment, the teachers all used some form of a four-point rubric to evaluate student understanding, but then recorded and combined the students' scores using different strategies. While there were some similarities between some of the teachers with respect to some practices, there were no two teachers who implemented standards-based grading practices in the same way across all aspects of the practice. As a consequence, the ability to easily compare a student's grade in one of the teacher's classroom to the grade of a student in a different teacher's classroom is decreased. This goes against the one of the arguments made by advocates of standards-based grading (e.g., Vatterott, 2015; Gentile & Lalley, 2003).

CHAPTER IX

DISCUSSION

The purpose of this qualitative, multicase study was to describe middle school mathematics teachers' teaching practices as they implemented stands-based grading. The results provide insights into the current literature on standards-based grading as well as informs current practice and future directions for investigation.

Connections To The Literature

The literature basis of the present study focused on how standards-based grading advocated that the evaluation practice should differ from traditional evaluation practices with respect to more than how teachers determine grades. Standards-based grading should change the focus of grades, the definition of grades, the structure of learning, and students' educational experiences (Vatterott, 2015). The four cases and cross-case analysis highlight challenges the teachers encountered with changing the (a) focus of grades, the (b) definition of grades, and the (c) structure of learning.

Focus Of Grades

Heflebower et al. (2014) and Vatterott (2015) claimed that grades should only reflect students' current understanding of content-specific knowledge and should reflect individual standards or learning targets (i.e., grade should not be a composite score). The teachers in this study differed in their (a) measurement of "current understanding," (b) communication of student grades, and (c) assessment of work habits.

Measuring current understanding. If a students' grade is meant to reflect a student's current understanding of content-specific material, then grades should only include measurements of that student's most recent attempt to demonstrate their understanding (Heflebower et al., 2014; Vatterott, 2015). That is, grades should focus students' understanding "at a particular point of time" (Guskey, 2009). Of the teacher participants, Mx. Johnson was the only teacher who modelled this idea in a way that advocates intended. Instead of including the students' highest score, Reilly included the students most recent score as part of the students' overall grade. As a result, it was possible for a student's grade to decrease if the student did not answer the question correctly. Mx. Johnson argued that the student was unable to retain the understanding and, therefore, no longer understood material; recording the students' new, lower score reflected that loss of understanding.

In contrast, the other three teachers used the students' highest grade on a standard, concept, or learning target. This is a variation of mastery-based grading (Gentile & Lalley, 2003) in which the teachers did not expect their students to reassess and demonstrate retention of understanding. For example, Mx. Williams' structure of concept quizzes expected students to complete a quiz for each concept until they achieved a passing score of three or four. Once the student achieved a passing score, Mx. Miller did not expect or require students to reassess on that concept. Students were not penalized for attempting to earn a higher score (i.e., trying for a four if they already earned a three). Mxs. Brown and Miller graded in similar ways. For the three teachers, instead of thinking of a student's grade as representing their current understanding as recommended by the literature (Guskey, 2009), the grades represented that these teachers' students achieved the understanding at some point during the course. As Mx. Williams' argued, the student's grade did not mean that the student currently understands the concept, but rather that the student understood the concept at one time and that the student might be able to relearn the material faster.

Communication and interpretation of grades. Gentile & Lalley (2003) recommended that, beyond measuring the students' current understanding, students' grades should reflect the students' understanding of specific content (i.e., standards or learning targets). Three of the four teachers recorded grades using a strategy that clearly document the specific content associated with that grade. Mxs. Johnson, Williams, and Miller recorded students grades by naming the associated items with the content those grades measured (i.e., competency, concept, and learning target, respectively). As a result, the literature (Gentile & Lalley, 2003) argues that it is more likely that stakeholders (e.g., students and parents) could use those grades to determine on which concepts a student is or is not proficient. Mx. Brown entered grades by naming items as the type of assessment (e.g., Chapter 16 Test). As a result, it is unlikely that a stakeholder could determine for which concepts a student is or is not proficient unless the stakeholder had additional knowledge of the assessment.

Deddeh et al. (2010) argued that composite scores that combine multiple measures of different concepts cause issues for communicating and interpreting students' grades. However, as reported in the case reports, the teachers' school districts required via policies and gradebook development that the teachers calculated a composite score. Mxs. Brown's, Johnson's, and Miller's school district-mandated gradebook reported grades with two levels of composite scores: (a) with respect to the standard, and (b) as an overall score. In contrast, Mx. Williams reported 30 individual scores, but the school district-mandated gradebook also reported an overall composite score. Even though three of the four teachers entered grades in a way that increased interpretability, the use of composite scores subsequently decreased interpretability.

Measurement of work habits. If students' grades should reflect their current content-specific knowledge, then the grades should not include measurements of non-content related criteria (i.e., work habits) (Gentile & Lalley, 2003; Vatterott, 2015). Deddeh and colleagues (2010) recommended measuring student behavior by creating a separate grade category called work habits. Mxs. Brown's, Johnson's, and Miller's school districts required measurement of specific work habits; however, Mx. Brown did not regularly measure their students' work habits. Mx. Williams' school district recommended that teachers assess their students' work habits but did not provide a formal list of habits or rubrics for assessment.

While the literature recommends creating a separate grade category, teachers should not include work habits scores in students' overall grade calculations because including those scores causes grades to no longer reflect student understanding of mathematical content (Deddeh et al., 2010) and it negatively impacts student motivation (Brookhart, 1994). Of the two teachers who regularly measured work habits, Mx. Johnson did not include work habits as part of their students' grades, but Mx. Miller included work habits as 1% of their students' overall grades. The two teachers used similar strategies for evaluating work habits, the difference in outcomes is a result of their school district-issued gradebooks. As is the recommendation from the literature (e.g., Deddeh et al., 2010), Mx. Johnson's gradebook reported two overall scores: (a) a content score and (b) a work habits score. Mx. Miller's gradebook did not provide the option for two reported scores.

Definition Of Grades

A condition of implementing standards-based grading with fidelity is to base evaluation on well-defined standards and criteria which outlines "the knowledge and skills students should have at each grade that prepares them for the next grade" (Bigham, 2015, p. 4). Mathematical standards suggested that "mathematical understanding and procedural skill are equally important, and both are assessable using mathematical tasks of sufficient richness" (NGA & CCSSO, 2010, p. 4). The teachers in this study measured students' knowledge and skills against a wide-range of "standards" and used varied levels of cognitive demand to teach and assess those standards.

What are "standards"? Each teacher's process for defining the standards against which they assessed students was very different, even though they taught in a state with mandated standards. Even in a context supportive of a common definition of standards across teachers and even school districts (i.e., the CAS-M standards), this research study found that defining standards for assessment was left, in three of the four cases, to the teacher to determine, and it was done inconsistently. Mxs. Brown and Miller aligned their grading criteria with the CCSS-M at varied levels of specificity, Mx. Johnons aligned them with the CAS-M using a district-mandated list of competencies, and Mx. Williams created their own list of CAS-M concepts. While the teachers' use of standards docuements differed, each of these standards documents defined a focus on both procedural fluency and conceptual undersatnding.

Enactment of standards. The expectation of teaching standards-based grading, where districts implement the CCSS-M and, by extension, the CAS-M, was that teachers enact "mathematics tasks of sufficient richness" (NGA & CCSSO, 2010, p. 4). Results for this study found that three of the four teachers used high-cognitive demand tasks on a regular basis. However, those teachers reported having to use external resources to supplement their district-issued curriculum to support instruction of high-cognitive demand tasks, and only two of the four teachers regularly included high-cognitive demand tasks on assessments.

Structure Of Learning

The aim of standards-based grading is to give all students the resources, including time, necessary to achieve a high-level of understanding. To achieve this aim, the structure of learning opportunities should include continuous formative assessment with feedback, differentiated instruction, and multiple opportunities to complete summative assessments (Gentile & Lalley, 2003; Heflebower et al., 2014; Vatterott, 2015). The degree to which these opportunities were present in the four teachers instructional design ranged from non-existent to present with limited fidelity.

Formative assessment and feedback. Three of the four teachers discussed the regular use of formative assessments but observations demonstrated that they did not implement the practice in a way that (a) coincided with the principles of standards-based grading (Guskey, 1997) or (b) provided feedback that improved student learning and teaching opportunities (Black & Wiliam, 1998).

Since formative assessments are meant to measure students' progression towards understanding, advocates for standards-based grading argued that teachers should not use formative assessment scores as part of grade calculations (Guskey, 1997); however, all three teachers who used formative assessment incorporated their students' formative assessment scores as part of overall grade calculations. In Mxs. Brown's and Miller's cases, the school district required the grade inclusion practice; whereas, Mx. Johnson used the scores as place holders until they administered summative assessments.

Instead of grading formative assessments, a teacher should use formative assessments as a communication tool between teacher and student focused on the student's understanding (Black & Wiliam, 1998). Only Mxs. Johnson and Miller used formative assessment as a two-way communication tool between themselves and their students. In Mx. Johnson's case, all students received written feedback; whereas, Mx. Miller's students only received feedback based on their performance, if the students indicated confusion or limited understanding via self-evaluation. Mx. Brown used formative assessments to inform their own view of student understanding but did not regularly give students feedback about that understanding.

In order to provide feedback to both teacher and student, advocates claim that formative assessments should be organized around well-defined learning targets and teachers should report results to students based on those learning targets (Vatterott, 2015). Only Mxs. Johnson and Miller communicated competencies and learning targets to their students as part of formative assessment. The results suggested that some of the teachers did not use their formative assessment practices to inform their students about specific learning targets for which the students had limited understanding.

Differentiated instruction. Advocates of standards-based grading argued that teachers should differentiate instruction in order to give students varied opportunities to learn as part of everyday instruction to account for the students' differences in understanding (Vatterott, 2015). Aligning with recommendations by Tomlinson and Eidson (2003), of the four teachers in this study, only Mx. Johnson engaged their students in regular differentiation of content, process, and product based on student readiness through their use of student-driven curriculum, strategic student seating, and projects that serve as alternatives to traditional summative assessments. Mxs. Brown, Williams, and Miller found it difficult to implement regular differentiation as part of classroom instruction due to limitations of classroom design (e.g., too many students and too few chairs) and curriculum; however, they argued that they made themselves available to students outside of class time. This justification suggested that the teachers viewed making themselves available to students as a part of or as a substitute for differentiated instruction.

Summative assessments. Vatterott (2015) recommended that teachers administer summative assessments to measure students' understanding after learning occurs organized around learning targets where they report scores for each target in the gradebook. Since students learn at different paces, advocates of standards-based grading recommended that students able obtain remediation and can reassess to improve their results (Vatterott, 2015). The views and practices of the teachers in this study with respect to summative reassessment ranged from no opportunities to endless opportunities. Mx. Miller did not allow their students to reassess because the time it would require was not practical within the constraints of their circumstances. However, Mxs. Brown, Johnson, and Williams were able to implement variations of reassessment by limiting students to one reassessment in Mx. Brown's case, building reassessment into regular assessment in Mx. Johnson's case, and by assessing using smaller and quicker assessments in Mx. Williams' case. The results supported the literature recommendation for summative reassessment within a mathematics classroom; however, it was unclear which of the three teachers' reassessment structures best supports students ongoing demonstration of understanding.

Recommendations

The teachers highlighted tensions and challenges that they encountered as part of their implementations of standards-based grading. There are several ways, informed by the evidence, that might help close the gap between recommendations in the literature and the practice of standards-based grading; including creation of well-defined standards, development of resources, and professional development.

Supportive Standards

Across the four teachers, the use of and alignment with the available standards documents varied greatly. Mxs. Brown and Miller aligned their grades with the CCSS-M to differing degrees, Mx. Johnson aligned their grades with the CAS-M and used district-developed competencies, and Mx. Williams created their own list of concepts based on the CAS-M. Such discrepancy in alignment undermined the argument for standards-based grading that students be graded based on the same benchmarks which results in grades that have increased interpretability and transferability.

Stiggins (2014) identified two problems related to the use of standards as a driving force in assessment: (a) there were too many standards to assess (i.e., they were too specific) or (b) the standards were defined too broadly. This claim was supported by the experiences of the teacher participants in this study. For example, Mx. Miller argued that aligning at the overarching standard level, as required by their school district, was too broad to result in meaningful communication of student understanding. In contrast, both Mxs. Brown and Williams noted that there were too many “standards” (i.e., CAS-M evidence outcomes) to assess in a meaningful and repeatable way. The teachers’ experiences highlighted the existence of the “problems” Siggins (2014) identified and highlighted the need for a standards document that is supportive of standards-based instruction.

To achieve the goal of supporting standards-based instruction, such a standards document would need to find a balance between having too many standards to measure and having too few standards that become overly broad (Stiggins, 2014). That is, the document would need to include a reasonable number of measurable standards that support the clear communication of student understanding. The standards document, CAS-M, given to teachers by their districts did not support this goal. Consider, for example, the first standard, Number Sense, Properties, and Operations, for eighth grade which was taught, in part, by Mx. Brown. See Figure 29 for an excerpt of this standard. The structure of the standards document allowed for the potential of evaluating student understanding at three different levels of specificity: (a) standard, (b) grade level expectation [GLE], and (c) evidence outcome.

The standard level was too broad and provided little detail about specific areas of student understanding, and although the GLE level was more specific, it masked some of the mathematical detail relating to concepts it was meant to represent (e.g., integer exponents, scientific notation). Evaluating students at either of these levels resulted in grades which were too broadly defined to effectively communicate student understanding of specific mathematical concepts as evidenced by the gradebooks of Mxs. Brown and Miller.

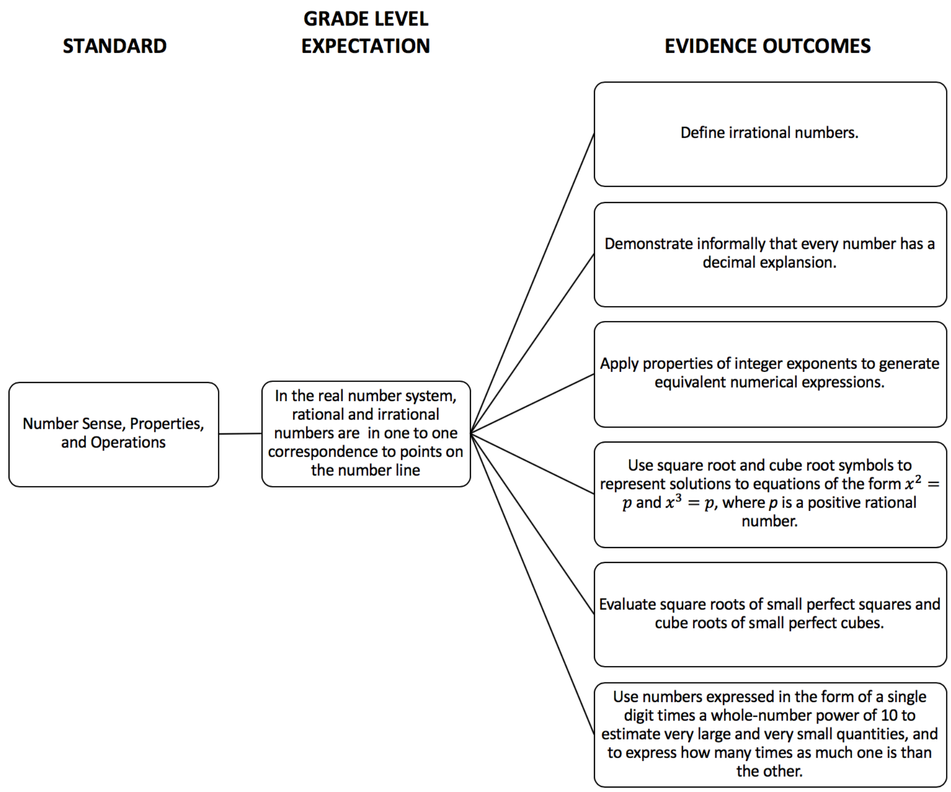

Figure 29. Graphical representation of an excerpt of the CAS-M first standard for Grade 8.

The evidence outcome level was too narrowly focused which resulted in the need for too many assessments. Mx. Williams' encountered this issue when developing their standards-based grading practices which resulted in the creation of their list of 30 assessable concepts. The experiences of the three teachers highlighted the limitations of the standards document as containing either too broad or too specific standards. This evidence supported a call for a revision to the standards document to support standards-based grading practices.

Resource Design

Having a standards document that supports the implementation of standards-based grading, teachers then need resources that continue to support their implementation of the practice (Knight & Cooper, 2019). The evidence suggested that the teachers' gradebooks and curriculum resources impacted their implementation of standards-based grading practices and resulted in the teachers needing to make choices about the extent to which they implemented the practices with fidelity. The evidence suggested the need for school districts to provide teachers with resources, including gradebooks and curriculum, that support the districts' mandated standards-based grading policies.

Gradebook. A gradebook that supports the principles of standards-based grading should allow for the reporting of student grades based on measurable standards without the inclusion of non-content related criteria and the use of composite scores (Gentile & Lalley, 2003). The four teachers in this research study all identified issues with the design and utilization of their district-issued gradebook. Such issues included problems with standards alignment, grade calculations, and grade conversions. To correct these issues, the teachers modified their gradebooks to some degree to fit within their perceptions of the goals of standards-based grading. The requirement of any modification to the gradebook increased the variability in teachers' use of the gradebook and introduced issues related to validity and reliability. School districts need to provide gradebooks that conform to the intensions of standards-based grading and do not require any modification to reduce the possibility of validity and reliability issues within and across teachers as such issues decrease interpretability of student grades.

Curriculum. If the intent of the standards is to encourage procedural and conceptual understanding, then teachers need to be given curriculum that supports instruction of procedural and conceptual understanding. For example, Mx. Johnson used such a curriculum and was regularly able to achieve high-cognitive demand instruction. In contrast, Mxs. Brown and Miller were given procedurally focused curricula. As a result, Mx. Miller felt they needed to spend additional time finding and incorporating conceptual understanding into their instructional practices. Mx. Brown, however, did not regularly engage in supplemental instruction focused on conceptual understanding and, instead, relied on the alignment of the curriculum to the standards. Having a curriculum that promoted both procedural and conceptual understanding is likely insufficient to ensuring that assessment and grades were aligned correctly (Knight & Cooper, 2019). As a result, such a curriculum should include explicit alignment with the standards. That is, it should be clear to both the teacher and student which measurable standard a curriculum resource or task is meant to teach and assess.

Professional Development

The teachers reported limited training and ongoing support with respect to their implementations of standards-based grading. Although the sample for the present study was small, statements by the participants indicated that their experiences (i.e., lack of district support) likely extended to their fellow teachers across their school districts. Training and support for teachers occur as they (a) complete their preservice teacher education program and (a) as ongoing, inservice teacher training activities (Battistone, Buckmiller, & Peters, 2019). Battistone et al. (2019) argued that "both teacher education programs and K-12 schools must take responsibility for better preparing and supporting early career teachers in their development in aspects of aligning curriculum, instruction, and assessment" (p. 15). While this claim is specific to early career teachers, it applies to the implementation of standards-based grading because, as a new practice, we can consider all teachers, preservice and inservice, to be new to the implementation of the practice.

Preservice teacher education programs. The teachers did not directly discuss their preservice teacher training programs with respect to their implementation of standards-based grading, but it was inferred from their perceived lack of prior training that their preservice teacher education programs did not directly train the teachers to implement the practice. Battistone et al. (2019) found that "the professors responsible for instructing assessment/grading theory influence pre-service educators' future behaviors by perpetuating traditional models of practice" (p. 15). As a result, "teacher educators need to model progressive assessment strategies" in their teacher education courses and help preservice teachers "make connections to assessment practices in the K-12 education system" (p. 15). That is, to effectively teach future teachers about standards-based grading practices, teacher educators have a responsibility to not only teach about, but actually use such practices in their own courses, including measuring against well-defined standards, implementing formative assessment, engaging in high-cognitive demand and differentiated instruction. As part of modeling grading practices, preservice teacher education programs also need to provide future teachers with opportunities to discuss, practice, and compare grading between expert and novice assessors before entering field placements (Grainger & Adie, 2014).

Inservice teacher support. School districts have a responsibility to provide and support effective, ongoing professional development related to assessment (Cizek, Fitzgerald, & Rachor, 1996). One way to support teachers in implementing standards-based grading is to provide ongoing professional development that ensures that teachers understand effective assessment practices and how to connect assessment to the standards and instructional practices. In a study of teachers' assessment practices, Cizek et al. (1996) found that, even though almost every school district had formal assessment policies, "only about one half of the teachers in [the] study said that they knew their district had a policy, and few of these teachers were able to supply any details about their districts' policies" (pp. 173-174). Like the results of Cizek and colleagues (1996), the evidence from the present study highlighted how the teachers' assessment strategies resulted in "a potpourri of elements that [varied] from district to district [and] from teacher to teacher within a district" (p. 174). This is highlighted by the evidence that no two teachers implemented standards-based grading in a similar way. That is, each teacher engaged in practices that were unique to their practice. The evidence suggested the need for ongoing professional development as well as well-defined district guidelines for the design and implementation of assessment strategies. To support this professional development, Stiggins (2014) recommended that school district personnel regularly ask themselves the following reflection questions to ensure that assessment practices become an effective part of teaching and learning:

> Are your learning targets clear and appropriate? Is your policy environment driving sound practice? Are your teachers [sic] assessment literate, and are they ready to communicate assessment results (to students, parents, school boards, administrators, and one another) in ways that support and certify achievement? (p. 75)

The questions imply that, as part of implementation, school districts should train teachers on assessment practices and make them ready to communicate the school district's goals and assessment mission to stakeholders.

DuFour, DuFour, Eaker, Many, and Mattos (2016) suggested collaborative professional learning communities (PLCs) as a strategy for supporting the ongoing professional development of inservice teachers. However, they cautioned that effective PLCs require "a school culture that is simultaneously loose and tight" (DuFour et al., 2016, p. 13). Such PLCs should be "loose" in the sense that administrators give teachers the freedom to implement the goals and resources developed by the group in a way that they believe best supports their individual students while still maintaining the spirit of the goals of the group. During initial implementation, Mx. Miller had the opportunity to engage in collaborative PLC work to develop resources and common practices. They cited this work as influential to their perceived success in implementing standards-based grading. In contrast, the PLC process needs to be "tight" in the sense that there are common elements and practices that are nonnegotiable with respect to achieving the overall goals of the group. One such elements is the idea that the PLC is a collaborative effort rather than a group of teachers working in isolation (DuFour et al., 2016). As a result, school leaders should require all teachers to participate and hold each other accountable for reaching the common goals with fidelity. By extension, this means that school districts have an obligation to facilitate regular common planning and discussion time for teachers either in common course teams or common content teams. All of the four teachers in this study reported either choosing not to or not have the ability to co-plan with other teachers. Another element is the mutual development and implementation of curriculum and assessment with a focus on supporting student learning (DuFour et al., 2016). Finally, DuFour and colleagues (2016) argued that effective PLCs must include the use of practice-based evidence to support meaningful reflection in order to improve individual as well as collaborative practice.

DuFour et al. (2016) claimed that PLCs themselves are not necessarily effective unless they are structured to achieve the goals in a particular way. That is, they argued that a "loose and tight culture will impact student and adult learning in a positive way only if the district is 'tight' on the right things" (p. 243). With respect to standards-based grading, the "right things" would include enforcing consistency in grading scale and grade calculation, improving the validity and reliability of assessments that support student learning, removing the inclusion of behavioral evaluation as part of grades, and developing resources to support students' in developing understanding (Knight & Cooper, 2019). Mx. Miller's original PLC work was specifically focused on analyzing and refining the school's implementation of standards-based grading within and across content areas. From their perspective, this allowed teachers to refine their practices and develop increased teacher buy-in and consistency in practice. In contrast, because of perceived decreased collaboration, Mx. Miller argued that there was increased inconsistency across teachers and less teacher buy-in.

Limitations And Future Directions

The research design and execution resulted in two key limitations: (a) restricted access to teachers for data collection, and (b) lack of data pertaining to the students' experiences of standards-based grading. These limitations open opportunities for additional research study of teachers' implementation of standards-based grading practices and the resulting experiences of their students.

Data Collection Access

The purpose of this multicase study was to describe the experiences of middle school mathematics teachers within and across school districts as they implement standards-based grading polices. As a result, a goal was to collect data from teachers from various school districts. However, access to possible teachers was limited due to requirements placed on the implementation of research studies by school districts. For some school districts, the deadline to apply for permission and access to conduct a research study was outside the feasible timeline for data collection for this study; whereas, another school district required the sharing of data (i.e., raw interview recordings) that raised concerns regarding research ethics and participants' right to anonymity. In several cases, requests for permission and access to conduct a research study were denied. In two cases, the school districts denied the existence of teachers who were implementing standards-based grading even though there existed evidence that such teachers did exist (Morgan & Powers, 2018). In another case, the school district denied the request because allowing the study might cause too much political conflict between the school district, parents, and teachers. Due to difficulties in obtaining access to potential teacher-participants, it is possible that the results presented here lack important variations that may have been uncovered from teachers within other school districts. That is, while the four teachers presented here offer qualitatively different practices, it could be that there is still more variation in practice that was not identified.

Future studies are needed that focus on exploring the practices of additional middle school teachers within and across different school districts. One such study might consider looking at several teachers within a single school district. Another study might focus on teachers with common characteristics including curriculum resources, grade level, and professional development. As part of identifying potential participants, future studies might consider identifying teachers who have been recognized as successfully implementing standards-based grading practices.

Student Experiences

Due to time and access constraints, collecting data on students' experiences as part of the implementation of standards-based grading was not possible, limiting the perspective to the experience of the classroom teacher. The literature claims that effective implementation of standards-based grading has the potential to impact and improve students' motivation to learn. Specifically, the practice might foster intrinsic motivation (Brookhart, 2011; McMillan, 2009), mastery goal orientations (McMillan, 2009), and growth mindsets (Boaler, 2016; Vatterott, 2015). Future research should incorporate measures of student motivation and attempt to document the students' perspectives and experiences with respect to the implementation of standards-based grading. As highlighted by the discussions of the results of this study, the teachers implemented standards-based grading in different ways. I hypothesize that the differing implementations had differing impacts on student experience and motivation. When attempting to determine which aspects of the teachers' practices were effective, it is important to know if those differences had a positive or negative impact on students' experiences.

Conclusion

The purpose of this qualitative, multicase study was to describe middle school mathematics teachers' teaching practices as they implemented stands-based grading. The evidence suggested qualitatively different practices as the result of the teachers' reconciliation of constraints within their school district and classroom. Individual case and cross-case analysis highlighted differences in the teachers' uses of mathematical tasks during instruction, implementation of instructional types, and teacher moves used to engage students in supporting student reasoning, assessment strategies, and evaluation practices. While there were some similarities among the teachers, there were no two teachers who implemented standards-based grading practices in the same way across all aspects of the practice. A consequence of this conclusion was that it was unlikely that a student's grade in one of the teacher's classroom was comparable to the grade of a different student in a different teacher's classroom. The evidence suggested the need for improved standards documentation, resource development, and professional development both at the preservice and inservice levels to better achieve the recommendations of the standards-based grading literature.

REFERENCES

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267. doi: 10.1037/0022-0663.80.3.260

Anderman, E. M., & Midgley, C. (2004). Changes in self-reported academic cheating across the transition from middle school to high school. *Contemporary Educational Psychology, 29*(4), 499-517. doi: 10.1016/j.cedpsych.2004.02.002

Anderman, E. M., Griesinger, T., & Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology, 90*(1), 84-93.

Austin, S., & McCann, R. (1992). *"Here's another arbitrary grade for your collection": A statewide study of grading policies*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Battistone, W., Buckmiller, T., & Peters, R. (2019). Assessing assessment literacy: Are new teachers prepared to assume jobs in school districts engaging in grading and assessment reform efforts? *Studies in Educational Evaluation*, *62*, 10-17. doi: 10.1016/j.stueduc.2019.04.009

Bigham, J. T. (2015). *The common core standards*. Indianapolis, IN: Penguin Group, Inc.

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. doi: 10.1007/s11092-008-9068-5

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8-21. doi: 10.1177/003172170408600105

Bloom, B. S. (1968). Learning for mastery: Instruction and curriculum. *Evaluation Comment, 1*(2), 1-12.

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. San Francisco, CA: Jossey-Bass.

Boesen, J., Helenius, O., Bergqvist, E., Bergqvist, T., Lithner, J., Palm, T., & Palmberg, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *The Journal of Mathematical Behavior, 33*, 72-87. doi: 10.1016/j.jmathb.2013.10.001

Bradbury-Bailey, M. (2011). A preliminary investigation into the effect of standards-based grading on the academic performance of African-American students. (Doctoral dissertation), Piedmont College, Demorest, GA.

Brodersen, R. M., & Randel, B. (2017). *Measuring student progress and teachers' assessment of student knowledge in a competency-based education system*. (REL 2017–238). Washington, DC: U.S. Department of Education, Institute of Education Services, National Center for Education Evaluation and Regional

Assistance, Regional Educational Laboratory Central. Retrieved from http://ies.ed.gov/ncee/edlabs.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35-36. doi: 10.1111/j.1745-39921991.tb00182.x

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

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

Brookhart, S. M. (2012). Grading. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257-271). Thousand Oaks, CA: SAGE Publications, Inc.

Burke, K. (2010). *Balanced assessment: From formative to summative*. Bloomington, IN: Solution Tree Press.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right, using it well* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Cizek, G. J., Fitzgeralds, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2), 159-179. doi: 10.1207/s15326977ea0302_3

Colorado Department of Education. (2010). *Colorado academic standards in mathematics*. Denver, CO: Department of Education.

Corey, S. M. (1930). Use of the normal curve as a basis for assigning grades in small classes. *School and society*, *31*(798), 514-516.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Crooks, A. D. (1933). Marks and marking systems: A digest. *The Journal of Educational Research*, *27*(4), 259-272.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge Grading: Endorsed by Students and Teachers Alike. *Applied Measurement in Education, 12*(1), 53-72. doi: 10.1207/s15324818ame1201_4

Crotty, M. (1998). *The foundation of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: SAGE Publications, Inc.

Danielson, C. (2008). Assessment for learning - Teachers as well as students. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 191-213). New York, NY: Lawrence Erlbaum Associates.

Davis, J. D. W. (1930). The effect of the 6-22-44-22-6 normal curve system on failures and grade values. *The Journal of Educational Psychology*, *22*(8), 636-640.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1-27. doi: 10.3102/00346543071001001

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

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*(2), 159-199. doi: 10.3102/00346543053002159

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23*(2), 167-180.

Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry, 14*(2), 129-149. doi: 10.1080/03626784.1984.11075917

DuFour, R., DuFour, R., Eaker, R., Many, T. W., & Mattos, M. (2016). *Learning by doing: A handbook for professional learning communities at work* (3rd ed.). Bloomington, IN: Solution Tree Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House, Inc.

Eells, W. C. (1930). An improvement in the theoretical basis of five point grading systems based on the normal probability curve. *The Journal of Educational Psychology*, *21*(2), 128-135.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*(3), 169-189. doi: 10.1207/s15326985ep3403_3

Elliot, A. J., & McGregor, H. A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501-519. doi: 10.1037/0022-3514.80.3.501

Ellis, A., Ozgur, Z., & Reiten, L. (2019). Teacher moves for supporting student reasoning. *Mathematics Education Research Journal*, *31*(2), 107-132. doi: 10.1007/s13394-018-0246-6

Engeström, Y. (2000). Activity theory and the social construction of knowledge: A story of four umpires. *Organization, 7*(2), 301-310. doi: 10.1177/135050840072006

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156. doi: 10.1080/13639080020028747

Engeström, Y. (2015). *Learning by expanding: An activity-theoretical approach to development research* (2nd ed.). New York, NY: Cambridge University Press.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review, 5*(1), 1-24. doi: 10.1016/j.edurev.2009.12.002

Fennell, F., Kobett, B., & Wray, J. A. (2015). Classroom-based formative assessments: Guiding teaching and learning. In C. Suurtamm (Ed.), *Annual perspectives in mathematics education: Assessment to enhance teaching and learning* (pp. 51-62). Reston, VA: National Council of Teachers of Mathematics.

Franklin, A. E. (2016). *Growth mindset development: Examining the impact of a standards-based grading model on middle school students' mindset characteristics.* (Doctoral dissertation), Drake University, Des Moines, IA.

Gentile, J. R., & Lalley, J. P. (2003). *Standards and mastery learning: Aligning teaching and assessment so all children can learn*. Thousand Oaks, CA: Corwin Press, Inc.

Grainger, P. R., & Adie, L. (2014). How do preservice teacher education students move from novice to expert assessors? *Australian Journal of Teacher Education*, *39*(7), 89-105. doi: 10.14221/ajte.2014v39n7.9

Guskey, T. R. (1996). Reporting on student learning: Lessons from the past – Prescriptions for the future. In T. R. Guskey (Ed.), *Communicating student learning: 1996 ASCD yearbook* (pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (1997). *Implementing Mastery Learning* (2nd ed.). Belmont, CA: Wadsworth Publishing Company.

Guskey, T. R. (2009). Grading policies that work against standards…and how to fix them. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 9-26). Thousand Oaks, CA: Corwin Press, Inc.

Guskey, T. R. (2011). Stability and change in high school grades. *NASSP Bulletin, 95*(2), 85-98. doi: 10.1177/0192636511409924

Guskey, T. R., Swan, G. M., & Jung, L. A. (2010). *Developing a statewide, standards-based student report card: A review of the Kentucky initiative*. Paper presented at the American Educational Research Association, Denver, CO. Conference Paper retrieved from http://eric.ed.gov/?id=ED509404

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

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

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and support. *Journal for Research in Mathematics Education, 28*(5), 524-549. doi: 10.2307/749690

Herbst, P., & Chazan, D. (2003). Exploring the practical rationality of mathematics teaching through conversations about videotaped episodes: The case of engaging students in proving. *For the Learning of Mathematics, 23*(1), 2-14.

Hochbein, C., & Pollio, M. (2016). Making grades more meaningful. *Phi Delta Kappan, 98*(3), 49-54.

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

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349-371. doi: 10.1037/0021-9010.64.4.349

Improving America's Schools Act, 103 U.S.C. §1111 (1994).

Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, *103*(1), 65-92. doi: 10.1177/0192636519826709

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience, 1*(2), 75-86. doi:10.1093/scan/nsl013

McMillan, J. H. (2009). Synthesis of issues and implications for practice. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 105-120). Thousand Oaks, CA: Corwin Press, Inc.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*(1), 487-503. doi:10.1146/annurev.psych.56.091103.070258

Merriam, S. B. (1995). What can you tell from an N of 1?: Issues of validity and reliability in qualitative research. *PAACE Journal of Lifelong Learning, 4*, 51-60.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementing*. San Francisco, CA: Jossey-Bass.

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

Morgan, M., & Powers, R. (2018). Traditional versus standards-based grading: A comparison of the teaching practices of secondary mathematics teachers. In T. E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), *Proceedings of the 40th annual meeting of the North American chapter of the international group for the psychology of mathematics education* (pp. 1146). Greenville, SC: University of South Carolina & Clemson University.

Murdock, T. B., & Anderman, E. M. (2006). Motivational Perspectives on Student Cheating: Toward an Integrated Model of Academic Dishonesty. *Educational Psychologist, 41*(3), 129-145. doi: 10.1207/s15326985ep4103_1

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Department of Education.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common core state standards. Washington, DC: National

Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from http://www.corestandards.org/Math/.

Nilson, L. B. (2015). *Specifications grading: Restoring rigor, motivating students, and saving faculty time*. Sterling, VA: Stylus Publishing, LLC.

No Child Left Behind Act, 107 U.S.C. §1111 (2001).

O'Connor, K. (2009). *How to grade for learning: K-12* (3rd ed.). Thousand Oaks, CA: Corwin Press, Inc.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

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

Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology, 103*(3), 683-700. doi: 10.1037/a0023911

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13. doi: 10.1002/bs.3830280103

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380. doi: 10.1016/j.tate.2010.03.008

Roth, W.-M. (2014). Activity theory in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 11-15). New York, NY: Springer.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67. doi: 10.1006/ceps.1999.1020

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology, 45*(4), 736-750. doi: 10.1037/0022-3514.45.4.736

Ryan, S., & Cox, J. D. (2017). Investigating student exposure to competency-based education. *Education Policy Analysis Archives, 25*(24), 1-32. doi: 10.14507/epaa.25.2792

Schwandt, T. A. (2007). *The SAGE dictionary of qualitative inquiry* (3rd ed.). Los Angeles, CA: SAGE Publications, Inc.

Smith, M. S., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3*(5), 344-350.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE Publications, Inc.

Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: The Guilford Press.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *The School Review*, *20*(17), 442-457.

Starch, D., & Elliott, E. C. (1913). Reliability of grading work in mathematics. *The School Review*, *21*(4), 254-259.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488. doi: 10.3102/00028312033002455

Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Stiggins, R. (2014). *Revolutionize assessment: Empower students, inspire learning*. Thousand Oaks, CA: Corwin Press, Inc.

Thorndike, E. L., & Bregman, E. O. (1924). On the form of distribution of intellect in the ninth grade. *The Journal of Educational Research*, *10*(4), 271-278.

Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum, 75*(3), 210-227. doi: 10.1080/00131725.2011.577669

Tomaz, V. S., & David, M. M. (2015). How students' everyday situations modify classroom mathematical activity: The case of water consumption. *Journal for Research in Mathematics Education, 46*(4), 455-496.

Tomlinson, C. A. (2008). The goals of differentiation. *Educational Leadership, 66*(3), 26-30.

Tomlinson, C. A. (2014a). The bridge between today's lesson and tomorrow's. *Educational Leadership, 71*(6), 10-14.

Tomlinson, C. A. (2014b). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A., & Eidson, C. C. (2003). *Differentiation in practice: A resource guide for differentiating curriculum*. Alexandria, VA: Association for Supervision and Curriculum Development.

Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wells, G. (1996). Using the tool-kit of discourse in the activity of learning and teaching. *Mind, Culture, and Activity, 3*(2), 74-101. doi: 10.1207/s15327884mca0302_2

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, Inc.

Wentzel, K. R., & Brophy, J. E. (2014). *Motivating students to learn* (4th ed.). New York, NY: Routledge.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). New York, NY: Lawrence Erlbaum Associates.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302-314. doi: 10.1080/00461520.2012.722805

APPENDIX A

INSTITUTIONAL REVIEW BOARD
APPROVAL LETTER

UNIVERSITY OF
NORTHERN COLORADO

Institutional Review Board

DATE:	October 12, 2017
TO:	Michelle Morgan
FROM:	University of Northern Colorado (UNCO) IRB
PROJECT TITLE:	[1122920-2] Standards-based grading practices in middle school mathematics classrooms: A multicase study
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	October 12, 2017
EXPIRATION DATE:	October 12, 2018
REVIEW TYPE:	Expedited Review

Thank you for your submission of Amendment/Modification materials for this project. The University of Northern Colorado (UNCO) IRB has APPROVED your submission. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on applicable federal regulations.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of October 12, 2018.

Please note that all research records must be retained for a minimum of three years after the completion of the project.

If you have any questions, please contact Sherry May at 970-351-1910 or Sherry.May@unco.edu. Please include your project title and reference number in all correspondence with this committee.

Michelle -

Thank you for providing clear and thorough amendments to your assent forms and the plans (included with the cover sheet) for obtaining school district approval(s). The first reviewer, Wendy Highby, has provided approval based on the modifications submitted. Subsequently, I've reviewed your original and modified materials and am also recommending approval.

Please use all amended materials in your participant recuitment and data collection. Best wishes with this meaningful research.

Sincerely,

Dr. Megan Stellino, UNC IRB Co-Chair

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within University of Northern Colorado (UNCO) IRB's records.

APPENDIX B

INITIAL INTERVIEW PROTOCOL

Introduction:

Thank you for agreeing to meet with me. I'm looking forward to hearing about your instructional practice and implementation of standards-based grading. The goal of this interview is to better understand your teaching practice and classroom norms. I'm going to ask you some general questions your teaching practice and your classroom norms. In addition, I'd like to ask you about the up-coming lessons that I'm going to be observing. Before we begin, what questions do you have?

Is there anything you'd like me to know about you, your class, or the up-coming lessons?

Teaching Practice Questions:

- How would you describe yourself as a mathematics teacher?
- What is you preferred teaching style? Why do you prefer this style?
- What does a typical mathematics lesson look like in your classroom?
- How do you plan for a unit?
- How do you plan for a lesson?

Classroom Norm Questions:

- How would you describe your class? Who are your students?
- Do you have any issues implementing your preferred teaching method with your students? Please explain.
- What challenges exist in this class?
- How do you handle issues with student motivation?
- How do you handle issues with student discipline?

Standards-Based Grading Questions:

- Was implementing this practice your choice?
    - If yes, why did you choose to implement this method of grading?
    - If not, what is your opinion of standards-based grading?
- Please describe the process of initial implementation.
- How would you describe your current implementation of standards-based grading?
- What course policies do you have in place to support this grading practice?
- Do you perceive any support for this grading method? Please explain.
- Do you perceive any resistance for this grading method? Please explain.
- What, if any resources, do you use?

Up-Coming Lesson Questions:

- Have you taught the up-coming lessons before?
    - If yes, how did they go the last time you taught them?
    - If not, how do you feel about teaching something new?
- What content do you plan to present in the up-coming lessons?
- What do you hope students will gain from the up-coming lessons?
- What types of learning opportunities do you have planned?
- Are you planning any assessments? If so, what are they?

APPENDIX C

SUMMARY AND REFLECTION GUIDING QUESTIONS

Guiding Questions

Lesson Plan Summary

Instructions:
Prior to teaching each lesson, take 5 to 10 minutes and, using the voice-recorder provided to you, summarize your up-coming lesson. Please use the questions below as a guide for completing this summary.

Guided Summary Questions:
- What are your goals for the up-coming lesson? Why are these goals important?
- How do you plan to implement the up-coming lesson?
- Why have you chosen to present the mathematical content in the way you have?
- How did you decide what to present (or not present) in the lesson?
- What (if any) standards will you be addressing today?
- Are you planning to conduct any assessments as part of today's lesson? If yes, what type of assessments will you implement, and what do you hope to learn from the assessments?

Lesson Plan Reflection

Instructions:
After teaching each lesson, take 5 to 10 minutes and, using the voice-recorder provided to you, reflection your lesson. Please use the questions below as a guide for completing this reflection.

Guided Reflection Questions:
- Did your lesson go as planned? Please explain.
- Did you implement any assessments as part of your lesson? If yes, how will you use the result of those assessments?
- Where there aspects of your lesson that you decided while instructing? If yes, what were they, and why did you make these decisions?
- Is there anything from this lesson that will inform or impact up-coming plans and lessons?

APPENDIX D

FINAL INTERVIEW PROTOCOL

Introduction:
Thank you again for agreeing to let me observe your class over the past week. The goal of this interview is to discuss the observed lessons as well as to better understand your teaching practice and classroom norms.
Before we begin, what questions do you have?
Before I get to my questions, is there anything you'd like me to know about the previous lessons?

Observed Lesson Questions:

- Would you say that the lessons I observed were typical of your class? Please explain.
- How do you think your lessons went? Please explain.
- Do you think you achieved your goals for this week? Please explain.
- Why do you think your lessons were (or were not) successful?
- Did your planned instructional activities go as planned? Please explain.
- Were there any instances of student thinking that surprised you? If yes, how do you think you handled those situations?
- If you were to teach these lessons again, how would you do it differently?

Standards-Based Grading Questions:

- Do you have any policies specifically implemented as part of your use of standards-based grading?
    - How were those policies developed?
    - Did you experience any resistance to implementing these policies?
- How often do you implement assessments as part of your class?
    - What is the purpose of these assessments?
    - Do you use the assessments as part of your lesson planning process? Please explain.
    - Do you expect your students to use the assessments as part of their learning process? Please explain.
    - Are students allowed to retake assessments?
- Do you give students feedback about their understanding of content?
    - When/how do you give this feedback?
    - Can you give an example of the type of feedback you might give students?
- How do you calculate student grades?
    - What type of evidence do you include in grade calculations?
    - Are students given an opportunity to change their grades?
- Do you offer opportunities for remediation?
- How do you determine which standards to address?

Concluding Questions:

- Do you have any advice for teachers who would like to implement standards-based grading in their classroom?
- Did participating in this research study change or impact your teaching practice? Please explain.